Nonlinear Stochastic Modelling of Antimicrobial resistance in Bacterial Populations

Philipsen, Kirsten Riber

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Nonlinear Stochastic Modelling of Antimicrobial resistance in Bacterial Populations

Kirsten Riber Philipsen
PhD Thesis

DTU Informatics
Technical University of Denmark

November 30, 2009

Supervisor
Henrik Madsen
Lasse Engbo Christiansen

Preface

This thesis was submitted at the Technical University of Denmark (DTU), department of Informatics and Mathematical Modelling (DTU Informatics) in partial fulfilment of the requirement for acquiring the PhD degree in engineering. The PhD project has been carried out as part of the project "Evolution and adaptation of antimicrobial resistance in bacterial populations" founded by the Danish Research Council for Technology and Production Sciences through the grant 274-05-0117.

The main collaborating institutions on the PhD project have been the University of Copenhagen, Department of International Health, Immunology and Microbiology, and Technical University of Denmark, DTU Food. Half a year of the PhD study have been carried out in Utrecht, the Netherlands in collaboration with the University Medical Center Utrecht, Department of Medical Microbiology, and Utrecht University, Department of Mathematics.

The topic of this thesis is application of stochastic differential equations and discrete stochastic models to bacterial growth, evolution and spread. The Thesis consist of a summarizing report and 7 research papers written during the PhD study. Two research papers have been published in international journals, one paper has been published as a research report at DTU Informatics and four research papers are either submitted or under preparation for submission.

Summary

This thesis applies mathematical modelling and statistical methods to investigate the dynamics and mechanisms of bacterial evolution. More specifically it is concerned with the evolution of antibiotic resistance in bacteria populations, which is an increasing problem for the treatment of infections in humans and animals. To prevent the evolution and spread of resistance, there is a need for further understanding of its dynamics.

A grey-box modelling approach based on stochastic differential equations is the main and innovative method applied to study bacterial systems in this thesis. Through the stochastic differential equation approach, knowledge of continuous dynamical systems can be combined with strong statistical methods. Hereby, important tools for model development, parameter estimation, and model validation are provided when in connection with data.

The data available for the model development consist mainly of optical density measurements of bacterial concentrations. At high cell densities the optical density measurements will be effected by shadow effects from the bacteria leading to an underestimation of the concentration. To circumvent this problem a exponential calibration curve has been applied for all the data. This new curve was found to perform the best calibration in a comparison with other earlier suggested curves.

In this thesis a new systematic framework for model improvement based on the grey-box modelling approach is proposed, and applied to find a model for bacterial growth in an environment with multiple substrates. Models based on stochastic differential equations are also used in studies of mutation and conjugation. Mutation and conjugation are important mechanisms for the development of resistance. Earlier models for conjugation have described systems where the

substrate is present in abundant amounts, but in this thesis a model for conjugation in exhaustible media has been proposed.
The role of mutators for bacterial evolution is another topic studied in this thesis. Mutators are characterized by having a high mutation rate and are believed to play an important role for the evolution of resistance. When growing under stressed conditions, such as in the presence of antibiotics, mutators are considered to have an advantages in comparison to non-mutators. This has been supported by a mathematical model for competing growth between a mutator and a non-mutator population. The growth rates of the two populations were initially compared by a maximum likelihood approach and the growth rates were found to be equal. Thereafter a model for the competing growth was developed. The models show that mutators will obtain a higher fitness by adapting faster to an environment with antibiotics than the non-mutators. In another study a new hypothesis for the long term role of mutator bacteria is tested. This model suggests that mutators can work as "genetic work stations", where multiple mutations occur and subsequently are transmitted to the non-mutator population by conjugation.

Another study in this thesis is concerned with the spread of colonization with resistant bacteria between patients in a hospital and people in the related catchment population. The resistance considered is extended-spectrum beta-lactamases, and it is the first time a model has been developed for the spread of this type of resistance. Different transfer mechanisms are studied and quantified with the model. Simulations of the model indicates that cross-transfer of resistance between patients is the most important mechanism of transfer.

The mathematical models developed in this thesis have helped to an improved understanding of the evolution and spread of resistance. They are thus a prime example of the strength of combining microbiology and experiments with modelling.

Resume

Denne afhandling anvender matematiske modeller og statistiske metoder til at undersøge dynamikken af og mekanismer for bakteriel evolution. Mere specifikt omhandler den evolution af antibiotika resistens i bakterie population, som er et stigende problem for behandlingen af infektioner i mennesker og dyr. For a hindre eller forebygge evolutionen og spredningen af resistens, er der et behov for yderligere undersøgelser af dennes dynamik.

En grå-boks modellerings fremgangsmåde baseret på stokastiske differentialligninger er den hovedsagelige anvendte og innovative metode benyttet til at studere bakterie systemer i denne afhandling. Viden om kontinuerte dynamiske systemer kan blive kombineret med stærke statiske metoder ved at benytte stokastiske differentialligninger. Herved bliver vigtige redskaber for model udvikling, parameter bestemmelse og model validering muliggjort i forbindelse med data.

Data tilgængeligt for modeludviklingen består hovedsageligt af optisk densitet målinger af bakterie koncentrationen. Ved høje densiteter af celler vil målingerne med optiske densitet blive påvirket af en skyggeeffekt fra bakterierne hvilket vil medføre en underbestemmelse af koncentratioen. For at undgå dette problem er en eksponentiel kalibrerings kurve blevet anvendt på alle datasæt. Det er fundet at denne kalibreringskurve giver den bedste kalibrering i sammenligning med andre tidligere foreslåede kurver.

Et nyt systematisk system for forbedringer af modeller baseret på grå-boks modellerings metoden er foreslået og anvendt til at finde en model for bakterievækst i et miljø med flere substrater. Modeller baseret på stokastiske differentialligninger er også anvendt til studier af mutation og konjugation. Mutation og konjugation er vigtige mekanismer for udviklingen af resistens. Tidligere modeller for konjugation har beskrevet systemer, hvor substrater er tilgængelige i rigelige mængder,

men i denne afhandling foreslår vi en model for konjugation i udtømmeligt medie.
Mutator bakteriers rolle i bakterieudvikling, er et andet emne der er undersøgt i denne afhandling. Mutatorer er karakteriseret ved en høj mutations rate og de antages at spille en vigtig rolle for udviklingen af resistens. Når mutatorer gror under stressende forhold, så som ved tilstedeværelsen af antibiotika, menes de at have en fordel i forhold til ikke-mutatorer. Dette er blevet understøttet af en matematisk model for konkurrerende vækst imellem en mutator og en ikke-mutator population. Vækstraten for de to populationer er først blevet sammenlignet med en maximum likelihood metode, hvorved det blev fundet at vækstraterne er ens. Herefter er en model for konkurrerende vækst blevet udviklet. Modellen viser at mutatorer vil opnå en højere fitness end ikke-mutatorer ved at tilpasse sig hurtigere til miljøet med antibiotika. I et andet studie bliver en ny hypotese for den langsigtede rolle for mutator bakterier testet. Resultaterne af modellen foreslår, at mutatorer kan virke som "genetiske arbejds stationer", hvor flere mutationer opstår og efterfølgende er overført til non-mutator populationen ved konjugation.

Et andet studie i denne afhandling omhandler spredningen af kolonisering med resistente bakterier imellem patienter på et hospital og det relaterede distrikt. Resistensen, der bliver betragtet, er extended-spectrum beta-lactamases, og det er første gang, at en model er udviklet for denne type af resistens. Forskellige overførselsmekanismer er studeret og kvantificeret med modellen. Simuleringer af modellen indikerer, at kryds-overførsel imellem patienter er den vigtigste mekanisme for overførsel.

De matematiske modeller udviklet i denne afhandling har medvirket til en forbedret forståelse af evolution og spredning af resistens. De er således et vigtigt eksempel på styrken ved at kombinere mikrobiologi og eksperimenter med modellering.

List of publications

The thesis is based on the following seven scientific research papers,

A Philipsen, K. R., Christiansen, L. E., Mandsberg, L. F., Hasman, H., Madsen, H., 2010. Comparison of calibration curves for the relation between optical density and viable cell count data. Submitted.

B Philipsen, K. R., Christiansen, L. E., Mandsberg, L. F., Ciofu, O., Madsen, H., 2008. Maximum likelihood based comparison of the specific growth rates for *P. aeruginosa* and four mutator strains. Journal of Microbiological Methods 75 (3), 551-557.

C Philipsen, K. R., Christiansen, L. E., Madsen, H., Modelling bacterial growth in rich media with a non-parametric extension to an SDE based model. Manuscript under preparation.

D Philipsen, K. R., Christiansen, L. E., Hasman, H., Madsen, H., 2010. Modelling conjugation with stochastic differential equations. Journal of Theoretical Biology 263 (1), 134-142.

E Philipsen, K. R., Christiansen, L. E., Mandsberg, L. F., Ciofu, O.,Madsen, H., Mathematical model for competitive growth of *P. aeruginosa* and four mutator strains in sub-MIC concentration of ciprofloxacin. Manuscript under preparation.

F Philipsen, K. R., Christiansen, L. E., Madsen, H.,Aarestrup, F. M., Mutators, a way to bypass classical Darwinism. Manuscript under preparation.

G Philipsen, K. R., Bootsma, M. C. J., Leverstein-van Hall, M. A., Cohen Stuart, J., Bonten, M. J. M., 2009. Dynamics of spread of intestinal colonization with extended-spectrum beta-lactamases in *E.coli*: a mathematical model IMM-Technical report-2009-13.

Other publications, which I have contributed to, but which will not be discussed further,

> Mandsberg, L. F., Maciá, M. D., Oliver, A., Philipsen, K. R., Christiansen, L. E., Alhede, M. , Kirkby, N., Høiby, N., Ciofu, O., 2010. Development of antibiotic resistance and up-regulation of the anti-mutator gene pfpI in mutator P. aeruginosa defective in two DNA Oxidative Repair Genes (mutYmutM). Submitted.
>
> Møller, J. K., Philipsen, K. R., Christiansen, L. E., Madsen, H., 2010. Analysis of diffusion in an SDE-based bacterial growth model. Submitted.

Acknowledgement

First, I would like to thank my supervisors Henrik Madsen and Lasse Engbo Christiansen for excellent support and fruitful discussions. I appreciate Henrik's positive mind and his encouragement, without which this thesis would not consist of such a high number of publications. I appreciate Lasse's assistance with planning of experiments and development of mathematical models as well as his help with everyday problems like improvement of simulation code.
I have enjoyed working on a cross-disciplinary project, and I would like to thank all my collaborators on the project. Especially, Henrik Hasmann for his work on conjugation, and his willingness to try out different ideas with his experimental studies. I would also like to thank Frank Aarestrup for interesting discussions concerning the evolution of bacteria. Finally I would like to thank Oane Ciofu and Lotte Mandsberg for many valuable discussions concerning the growth and evolution of *Pseudomonas aeruginosa*, and the possibility to use their data for my study.
I am grateful to Marc Bonten for making my stay at the University Medical Center Utrecht (UMCU) in the Netherlands possible. I would like to thank Marc Bonten, Maurine Leverstein-van Hall and James Cohen Stuart from the UMCU and Martin Bootsma from the Department of Mathematics at Utrecht University for many valuable discussions and their comments on my work.
Furthermore, I would like to thank my colleagues at IMM Informatics for making the last three years of work an enjoyable time. Especially thanks to the colleagues helping me out with smaller and bigger questions related to e.g. statistics, stochastic processes and `R` programming. Also a special thanks to the lunch club for many interesting conversations over lunch and to the Friday bar group for some nice Friday afternoons.
Finally, I would like to thank my boyfriend Raymond Bergmann, for his continuous support and believe in that I would make it, and for his honest and constructive feed-back on my work.

List of abbreviations

AIC	Akaike's Information Criterion
BHI	Brain Heart Infusion
CF	Cystic Fibrosis
CFU	Colony Forming Unit
CTSM	Continuous Time Stochastic Modelling
EKF	Extended Kalman Filter
ESBL	Extended-spectrum Beta-lactamases
FDSA	Finite Difference Stochastic Approximation
ICU	Intensive Care Unit
LB	Luria-Bertani
LS	Least squares
MAP	Maximum a Posteriori
MIC	Minimum Inhibitory Concentration
ML	Maximum Likelihood
MRSA	Methicillin-resistant *Staphylococcus aureus*
OD	Optical Density
ODE	Ordinary Differential Equation
SA	Stochastic Approximation
SDE	Stochastic Differential Equation
SPSA	Simultaneous Perturbation Stochastic Approximation

Contents

CHAPTER

1 Introduction

Treatment of tonsillitis, recovery of skin infections and prevention of infections during surgery is a few examples of areas, where antibiotics play a major role. Antibiotics are developed to combat "bad" bacteria populations. It should be remembered that not all bacteria population are bad. In our intestine for example bacteria live, which are essential for the digestion of food. The problem arises if the bacteria get establish outside its natural habitat, such as intestinal bacteria causing urinary tract infections. Diseases caused by bacteria can be treated with antibiotics, as long as the bacteria are not resistant against the antibiotic. Unfortunately antibiotic resistance is increasing faster than new antibiotics are developed. This is a major concern as it leads to failure of treatment, prolonged illness, and in some cases death. Therefore the need for new treatment strategies or ways of slowing down the evolution of resistance are high. The aim of this thesis is to improve the understanding of the evolution process and spread of resistance, and herewith be part of changing the longer-term effects of resistant bacteria.
New dynamical stochastic models have been developed for this thesis to support the above goal. Stochastic continuous time models in the form of Stochastic Differential Equations (SDEs) have to my knowledge only had very limited application within bacterial evolution. In the following chapters and papers importance and advantages of using SDEs is demonstrated in relation to modelling of bacterial evolution. Discrete stochastic models have been considered for more conceptual models for evolution and spread of resistance, where the number of bacteria or people are so low that it is not reasonable to use a continuous sample space.
Apart from this introduction the thesis consists of seven papers, and two summarizing chapters explaining important theory and placing the contributions in the papers in relation to each other and existing literature. In Chapter 2 the focus is on the mathematical technics used and developed for this thesis. The theory behind SDEs is introduced and the simulation of discrete stochastic models is described and discussed. In Chapter 3 mechanisms for growth and evolution of bacterial

populations are described, and the models developed in the papers are summarized. Furthermore a common conclusion is given in Chapter 4. The papers are ordered such that studies which form a mathematical basis for other papers are placed before these.

CHAPTER

2 Modelling of stochastic systems

Stochastic Differential Equations (SDEs) are widely used as a mathematical model for describing the dynamics of stochastic systems. SDEs are applied in a verity of scientific fields such as insulin secretion (Mortensen et al., 2007), pharmacokinetics and pharmacodynamics (Klim et al., 2009), the geolocation of fish (Pedersen et al., 2008), heat dynamics of buildings (Madsen and Holst, 1995; Andersen et al., 2000; Jimenez et al., 2008), water flow in a river (Jacobsen et al., 1997; Jonsdottir et al., 2001) and stock prices (Nielsen et al., 2000). Similar to an Ordinary Differential Equation (ODE), an SDE is a continuous time, continuous state model. However, while the solution to an ODE is a deterministic function, the solution to an SDEs is a stochastic process. Thus, with an ODE description of a system, it is assumed that knowing the given state, we can exactly predict what will happen also in the far future. By modelling a system using SDEs, we recognize that this is not possible. In stead the SDE provide information about the systems stochastic behavior (measured by e.g. its mean, variance, and covariance), which enables the use of strong statistical tools for model development and validation.
The objective of this thesis is to model bacterial evolution while taking into account its stochastic nature. Bacteria populations are highly complex, and models used to describe growth and evolution of bacteria will therefore be a simplification of the real system, meaning that only important processes are included in the model. When using SDEs to model the dynamics of bacterial evolution noise is added to the deterministic part of the model, hereby accounting for those processes not included in the model. Before explaining some of the theory of SDEs an example of a typical data set for the growth of bacteria is given. The size, N_t, of a bacterial population during exponential growth can be modelled as

$$\frac{dN_t}{dt} = a_t N_t \ , \tag{2.1}$$

where a_t is the growth rate and the subscript t indicate a time dependence of N and a. The growth rate might change slightly over time due to some random ef-

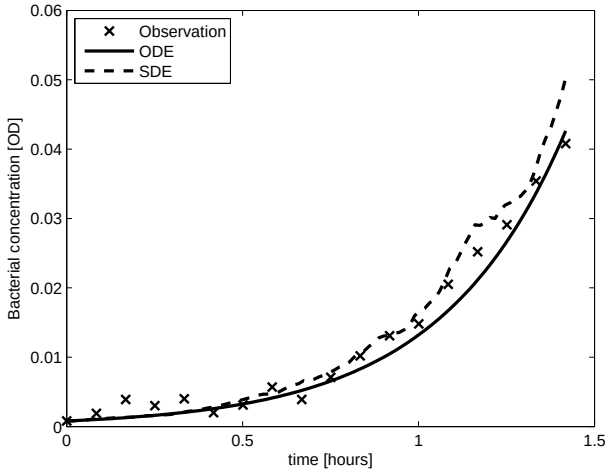


Figure 2.1: The growth of Pseudomonas aeruginosa is an example of a typical dataset for which modelling with SDEs is required to obtain an accurate parameter estimation.

fects, e.g. change of temperature, oxygen availability and pH, which are not included in the model. In that case the growth rate is

$$a = \mu + \text{"noise"}\,, \tag{2.2}$$

where μ is non-random. Once noise is added to μ, the bacteria concentration is a random process, which means that N no longer has an unique value but can be described e.g. by its moments. Figure 2.1 shows the concentration of *Pseudomonas aeruginosa* obtained from optical density (OD) measurements plotted together with the ODE model and one realization of the SDE model. The data layout shown in Figure 2.1 is typical. The observations do not fluctuate randomly around the model but are repeatedly located to one side. This leads to autocorrelated residuals when fitting the model, which can be described by introducing SDEs for describing the dynamics. The use of SDEs is especially important as it facilitates the use of prediction error methods for parameter estimation, as opposed to traditional output error (or simulation) methods used to estimate parameters in ODEs. An SDE combined with prediction error methods will provide estimates, which are closer to the real values. For the example considered here the growth rate is estimated to 2.80 using an ODE model whereas the growth rate for the SDE model is estimated to 2.66. There is a clear difference between the two estimates, and the SDE model gives a statistical better fit to the data. In order to understand the methods used for estimating parameters in an SDE, the theory of SDEs will be introduced in the following sections.

2.1 Diffusion

When adding noise to an else deterministic system it is desirable that the noise resembles the observed physical noise. An example of a physical system influenced by noise is paint being dropped in a well controlled stream of water. The flow of the drop of paint, along the stream can be described deterministically from physical laws, this is also called the drift of the system. However, the paint will not only flow along the stream but it will also diffuse outwards in the water. This diffusion process is caused by the noise present in the system. Due to diffusion the exact location of the paint at a later time can not be determined, but the probability distribution can. On a microscopic scale the diffusion of particles corresponds to each particle performing a random walk, or said in another way the movement of the paint particles is a stochastic process. A stochastic process in turn is a parameterized collection of random variables $\{X_t\}_{t\in T}$ defined on a probability space and assuming values in $\Re$. The probability space is described by the sample space Ω of possible outcomes, a σ-algebra $\mathscr{F}$ of subsets (events) of Ω and a probability measure P which assign a probability to each event. For a fixed time $t \in T$ the stochastic process reduces to a random variable whereas fixing $\omega \in \Omega$ gives a realization of X_t, which is also called a path. In Figure 2.2 three paths of the stochastic process X_t are shown, and three outcomes of the random variable for one specific time point are marked.

When considering many particles (such as the drop of paint) their location at a

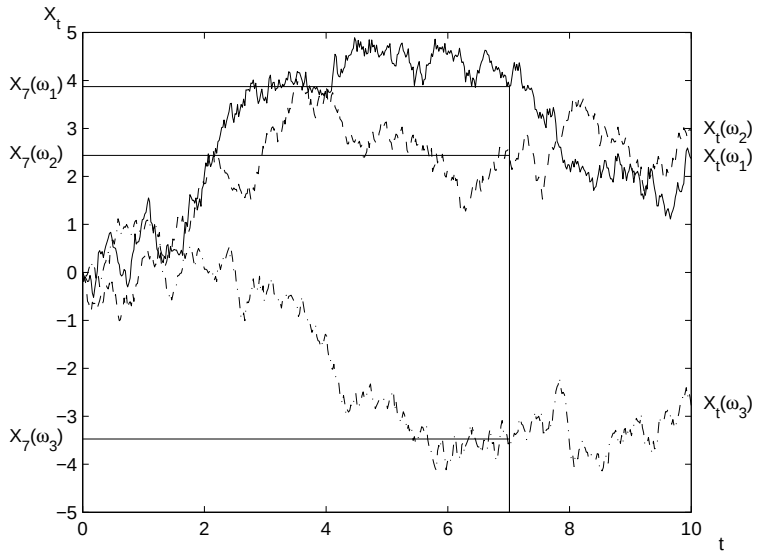


Figure 2.2: This figure shows three path of a stochastic process (which is in fact a standard Brownian motion) for $t \in [0,10]$. Each sample path represents a realization of the stochastic process X_t.

given time will be approximately normally distributed due to the central limit theorem. An important generalization of the random walk is, when the steps may happen at any time t and when the size of the steps follow the normal distribution. This process can be constructed by considering independent steps of length $\pm\sqrt{\Delta t}$ taken at time intervals of length Δt. Letting z_i be the step at time i then $P(z_i = -1) = P(z_i = 1) = 1/2$. Considering the process

$$Y_t = \sqrt{\Delta t}(z_1 + \cdots + z_{[t/\Delta t]}) \,, \tag{2.3}$$

where $[t/\Delta t]$ is the integer part of $t/\Delta t$, then the expected value $E[Y_t] \to 0$ and variance $V[Y_t] = \Delta t[t/\Delta t] \to t$ as $\Delta t \to 0$. Furthermore, the central limit theorem implies that $Y_t \to W_t$, where W_t is a Wiener process. A Wiener process beginning at 0, with drift μ and variance σ is a process which has the following properties

1. $P\{W_0 = 0\} = 1$.
2. For all non-overlapping time intervals $[t_1, t_2]$, $[t_3, t_4]$ the random variables $W_{t_2} - W_{t_1}$ and $W_{t_4} - W_{t_3}$ are independent.
3. For any time interval $[t_1, t_2]$, $W_{t_2} - W_{t_1}$ is gaussian distributed with

$$E[W_{t_2} - W_{t_1}] = \mu(t_2 - t_1),$$
$$V[W_{t_2} - W_{t_1}] = \sigma^2(t_2 - t_1).$$

If $\mu = 0$ i.e., the mean is 0 then the process is called a Brownian motion. If in addition $\sigma^2 = 1$ the process is called standard Brownian motion or a standard Wiener process. Thus, the increment that a standard Wiener process makes over a time interval $t_2 - t_1$ is normally distributed with mean 0 and variance $t_2 - t_1$. Note that the Wiener process is not stationary since $V[W_t] \neq V[W_s]$ if $s \neq t$; but the increment $W_{t+h} - W_t$ is stationary. Furthermore the Wiener process is nowhere differentiable (Øksendal, 2007).

2.2 Itô integrals

The reason for considering the diffusion process in the previous section is to be able to add noise to an ODE i.e. to describe equations of the form

$$\frac{dX_t}{dt} = f(t, X_t) + \sigma(t, X_t)\text{"noise"} \,, \tag{2.4}$$

where f and σ are some given functions. It is reasonable to look for some stochastic process β_t to represent the noise term such that

$$\frac{dX_t}{dt} = f(t, X_t) + \sigma(t, X_t)\beta_t. \tag{2.5}$$

Based on many physical situations, one is led to believe that β_t has, at least approximately these properties (Øksendal, 2007)

1. $t_1 \neq t_2 \Rightarrow \beta_{t_1}$ and β_{t_2} are independent,
2. β_t is stationary,
3. $E[\beta_t] = 0$ for all t.

It turns out that there does not exist any "reasonable" stochastic process satisfying these properties (Øksendal, 2007). The process would be a white noise process, which is possible in the discrete time case but a mathematical abstraction in the continuous case. According to the first property the process at two different time points could be very far apart, but this contradicts with the property of stationarity. A more reasonable construction is obtained by letting $0 = t_0 < t_1 < \cdots < t_m = t$ and considering a discrete version of (2.5)

$$X_{k+1} - X_k = f(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)\beta_k \Delta t_k \,. \tag{2.6}$$

We now replace $\beta_k \Delta t_k$ by $\Delta V_k = V_{k+1} - V_k$ where $\{V_t\}_{t\geq 0}$ is some suitable stochastic process. The assumptions 1, 2, and 3 on β_t suggest that V_t should have stationary independent increments with mean 0. Such a process exists and it turns out that the only such process with continuous paths is the standard Wiener process W_t. Thus we put $V_t = W_t$ and from (2.6)

$$X_k = X_0 + \sum_{j=0}^{k-1} f(t_k, X_k)\Delta t_k + \sum_{j=0}^{k-1} \sigma(t_k, X_k)\Delta W_j. \tag{2.7}$$

The limit of the first sum on the right hand side can be interpreted using e.g. the Riemann sums, but the interpretation of the second sum is not straight forward. In fact the value of the second sum depends on where in the interval $[t_j, t_{j+1}]$ it is evaluated. The following two choices have turned out to be the most useful ones

1. The **Itô integral** where the sum is evaluated at $t = t_j$ for $t \in [t_j, t_{j+1}]$, i.e. the left endpoint.
2. The **Stratonovich integral** where the sum is evaluated at $t = (t_j + t_{j+1})/2$ for $t \in [t_j, t_{j+1}]$, i.e. the midpoint of the interval.

In the remainder of this thesis the Itô integral will be used, as this has some advantages when used for parameter estimation. Thus, the stochastic differential equation

$$dX_t = f(t, X_t)dt + \sigma(t, X_t)dW_t \,, \tag{2.8}$$

should be understood as a process X_t, which is represented by the Itô integral

$$X_t = X_0 + \int_0^t f(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \,. \tag{2.9}$$

An important property of the Itô integral is that it is a Martingale, which means that the expected value of a process only depends on the present value and not on future values. Furthermore it is a Markov process, as future values depends only on the present value and not on past values.

2.3 Grey box modelling

If the whole dynamics of a system and the corresponding parameter values are know, a deterministic model such as an ODE would be a good choice for modelling the system. This is known as a white box model. In that case the model will be explicitly based on physical knowledge about the system. This also implies that the values of the states of the system can be predicted exactly for all future time points. If the system is either too complicated to describe by physical equations, or no physical knowledge is required to fulfil the purpose of the model, then a model based purely on data, such as an autoregressive model or neural network, is a better choice. This is called black-box modelling as no information about the underlying system is used. In this thesis the best of each of the two modelling approaches are combined in the SDE framework resulting in so-called grey box modelling. In grey-box modelling physical knowledge of a system is combined with data and statistical methods, to get the best description of the system. If all the states in the SDE could be measured directly without noise, then these observations could be used to estimate the parameters in the model directly. However, observations of states are almost always encumbered with noise, which should be separated from the system noise incorporated in the SDE. Therefore the following continuous discrete time stochastic state space model is used (Kristensen et al., 2004b)

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t\,, \tag{2.10}$$

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k\,, \tag{2.11}$$

where Equation (C.1) is the continuous time system equation (the Itô SDE) and Equation (C.2) is the discrete time observation equation. The notation has been extended in comparison to the previous section, to make it clear that $\boldsymbol{f}$ and $\boldsymbol{\sigma}$ depends on the model parameters $\boldsymbol{\theta}$ and a vector of known input variables, $\boldsymbol{u}_t$. In the observation equation $\boldsymbol{e}_k$ is an l-dimensional white noise process with $\boldsymbol{e} \in \mathcal{N}(0, \boldsymbol{R}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}))$, where $\boldsymbol{R}$ is the covariance matrix for the noise. The first term of (C.1) is called the drift term and the second term is called the diffusion term, which can intuitively be understood from the example of paint in water previously mentioned.
The unknown model parameters can be found by maximizing the likelihood function

$$L(\boldsymbol{\theta}; \mathscr{Y}_N) = \left(\prod_{k=1}^{N} p(\boldsymbol{Y}_k | \mathscr{Y}_{k-1}, \boldsymbol{\theta}) \right) p(\boldsymbol{Y}_0 | \boldsymbol{\theta}) \tag{2.12}$$

with respect to a known sequence of observations

$$\mathscr{Y}_k = [\boldsymbol{Y}_k, \boldsymbol{Y}_{k-1}, \ldots, \boldsymbol{Y}_1, \boldsymbol{Y}_0]\,. \tag{2.13}$$

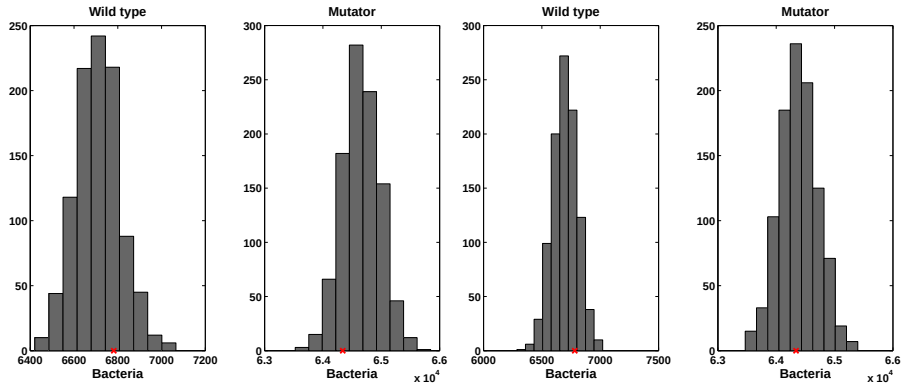
An exact evaluation of the likelihood function would require knowing the initial probability $p(\boldsymbol{Y}_0|\boldsymbol{\theta})$ and determining the subsequent conditional probabilities by

successively solving Kolmogorov's forward equation and using Bayes' theorem, but this approach is not computational feasible in practice (Kristensen et al., 2004b). Provided that the sample intervals are reasonable compared to the non-linear dynamics of the system, the conditional probability densities p can be approximated by Gaussian densities, as the SDE is driven by a Wiener process which has Gaussian increments. This enables the use of filtering techniques to determine the conditional probability densities. In this thesis the model parameters have been estimated by optimizing the likelihood function using an Extended Kalman Filter (EKF) approach as described in Paper C. The software `CTSM`* (Kristensen et al., 2004b) has been used to obtain the maximum likelihood estimates of the modelling parameters. This software uses the Extended Kalman filter to compute the log-likelihood for non-linear models and a quasi-Newton method to optimize the log-likelihood function. `CTSM` is used, because it has been shown to perform well in comparison to other methods (Kristensen et al., 2004b), and (maybe more important) it is easy to use and provides the optimized log-likelihood value, correlation between parameters and uncertainties on parameter estimates.

2.4 Discrete state stochastic models

For some systems the events to be modelled are rare and the time to first event important. It is therefore necessary to consider integer bacteria numbers in stead of the concentration of bacteria as is described in SDEs. The number of bacteria in a population should then be simulated as a discrete stochastic process. In continuous time the Poisson process can often be used. The probability of, e.g., a mutation in a bacterial population can be assumed to be independent of the history of the population; it depends only on the current population size and the mutation rate. In Paper F a discrete stochastic model for the evolution of bacteria is simulated using a fixed-increment time advance method. This means that for each fixed time interval the number of events that has occurred during that interval are sampled from a Poisson distribution (Law and Kelton, 2000). The events occurring during the time interval are all considered to take place at the end of the interval. This also means that the effect of an event on the system can not be effectuated before the end of the interval. If the time interval is too large the result can be, for example that a bacteria divide and die within the same interval.
Another simulation approach often used for queueing systems is discrete event simulation. In discrete event simulations the time until the next event and the type of event is first found, and hereafter, this event is performed which involves random number generation. The numbers could be generated for instance from a Poisson distribution. The discrete event simulation approach was implemented for the model in Paper F, in order to compare with the simpler fixed-increment time advance method. The two methods were found to give similar outcomes as seen in Figure 2.3, which shows the amount of bacteria in two different states af-

*www.imm.dtu.dk/~ctsm



(a) time step of 0.1 (b) time step of 0.01

Figure 2.3: Comparison of the number of bacteria with one mutation in each of the two states: wild-type and mutator. The histogram is the result of 1000 simulations using the fixed-increment time advance method and the red cross is the result of one run of the discrete event simulation. The p-values for the hypothesis that the two methods gives the same outcome is for a time step of 0.1 h: 0.48 for the wild-type and 0.33 for the mutator, and for a time step of 0.01 h: 0.43 for the wild-type and 0.90 for the mutator.

ter one hour of simulations. To perform the discrete event simulation, however, $1.4 \cdot 10^{12}$ events most be computed to simulate dynamics of the model for a single model. Continuing the discrete event simulation until five days would therefore be computationally infeasible due to the high number of events. The smaller the time step the more alike are the two simulation methods. The time step for the fixed-increment time advance approach was set to 0.1 h, which is around 1/6 of the growth rate. The computational time for each simulation is low using this time step, and it can not be rejected that the two methods perform equal outcomes (p = 0.48 for the wild-type and p = 0.33 for the mutator).

In Paper G a model for the spread of resistant bacteria between patients in a hospital and the catchment area is considered, including also the movement of patients to and from the hospital. Also here a fixed-increment time advance approach is used. The time step is set to one day, which is chosen as it is reasonable to assume that discharge from and admittance to the hospital happens once a day. With regards to the transfer of resistance this is a continuous process indicating that discrete event simulation should be the method of choice. However, the colonization of a person and further growth of the bacteria to reach concentrations high enough for the bacteria to be transferred to other people is not an instant process. A time step of one day would require that this process takes one day, which is judged to be reasonable. To ensure that two event do not happen to the same person, e.g. movement to another hospital ward and discharge, the events are sampled recursively from a multinomial distribution.

Markov chain simulation have also been considered for both studies. A Markov process is a process for which a future value of the process is only dependent on the present information about the process. For each iteration of the Markov chain (or Markov jump process in discrete time) a future state would depend on the current state and the probability of moving to each of the possible states (including staying in the current state). If the state space is finite the transition probabilities can be represented by a transition matrix. When modelling bacteria populations the state space is finite, but multivariate and very large as a bacterial population can grow from one bacteria to billions. Additionally the process is often non-linear as the rate of e.g. conjugation depends on the number of bacteria in two populations. Modelling the population dynamics as a Markov chain would therefore be computational infeasible.

2.5 Systematic model improvement

The use of SDEs in comparison to ODEs paves the way for several statistical methods of systematic model improvement (Kristensen et al., 2004a). The necessity of model improvement can be revealed from the SDE framework if, e.g. the correlation between two parameters is very high, or a high standard deviation is seen for the incremental covariance of the system noise of a given state. In this way model deficiencies can be revealed, in a way which cannot be facilitated by a deterministic model.
Paper C uses these ideas to introduce a new strong framework for systematic model improvement, and successfully apply it to find a better model for bacterial growth on rich media. The key element of the framework is SDEs, and it contains a parametric and a non-parametric part. The non-parametric extension to the model can be used to dig into embedded information of the dynamics of the system. The parametric part contains steps of model specification, parameter estimation, and validation.
Model validation may consist of a variety of statistical tests. One way of validating a model is to compare it with another model containing more parameters to describe the system. In this thesis several inference studies have been performed to examine which of two models should be chosen to describe a given system (Paper A, B, C, and D). For all studies a Maximum Likelihood (ML) based approach has been used to test the models. When comparing nested models a likelihood-ratio test is used to find the best model (Paper D and B). This approach is preferred compared to Akaike's Information Criterion (AIC), as the latter does not give an estimate for how much better one model is compared to the other. However, for un-nested models the AIC approach is a good alternative and it has been used in Paper A and C. In Paper B it was shown that the choice of parameterizations of the covariance matrix for the residuals of a regression model can influence the outcome of the test. Here the ML approach provides an advantage as it enables an simultaneous estimation of the parameters of the covariance matrix and the

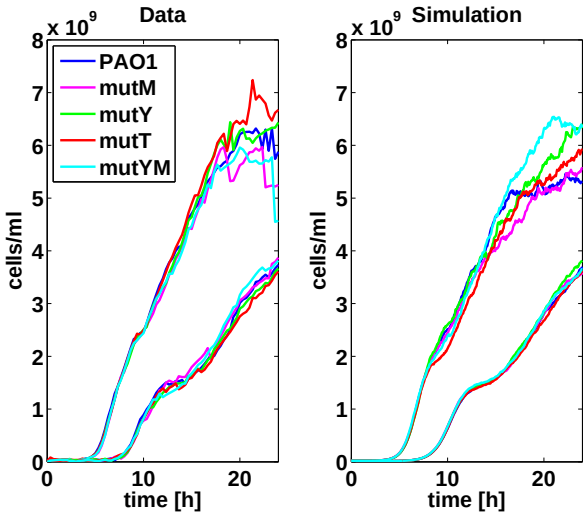


Figure 2.4: Data (left) and simulation result (right) for the concentration of *Pseudomonas aeruginosa* (PAO1) and the four PAO1 mutators *mutM*, *mutY*, *mutT*, and *mutY-mutM* when growing in Luria-Bertani media (top lines) and Luria-Bertani media with 0.1 μg/ml ciprofloxacin (bottom lines).

model parameters.

2.6 Discussion

`CTSM` has been found to be a very useful program for performing the analysis and estimations necessary for this thesis. The Kalman approach implemented in `CTSM` does, however, give one disadvantage: it is not possible to include state dependent noise in the system equation (Equation C.1), e.g. multiplicative noise. For some simple models the problem with multiplicative noise can be solved by performing a log-transformation of the states. For more complicated models another approach can be taken, as has been done in Paper C, D and E. This approach is to use the measured state variable instead of the state itself. Hereby a scaling for the standard deviation of the system noise is obtained which corresponds to multiplicative noise.How well this "trick" resembles a system with multiplicative noise has not been tested in the papers. It is only for the parameter estimation that multiplicative noise can not be used directly, whereas it can be implemented afterwards when running the simulations. An example is given in Figure 2.4(modified from Paper E), where a model with multiplicative noise is simulated with parameters estimated in `CTSM`. It is seen from the figure that the simulation very well captures the variation in the data. This indicates that using observed state values for a scaling of the system noise is a good alternative for multiplicative noise. It should be noted, however, that several other methods of evaluating the likelihood function exist (Ionides et al., 2006; Lele et al., 2007; Frydendall, 2009), for which

multiplicative noise can be used, but these have not been considered here.

The biological systems considered in this thesis have been modelled either with continuous state or discrete state models. In some situations a combination of the two might give a more correct representation of the physical system. For example in the study modelling the competition between two bacteria with different mutation rates (Paper E), where the time of occurrence of the first (whole) bacteria can have a large effect of the dynamics. A suggestion for an improvement of the SDE framework is therefore to incorporate a combination of discrete and continuous states in the parameter estimation. This would allow the method to be used on a wider class of problems.

CHAPTER

3 Bacterial evolution and spread

This chapter seeks to give a conceptual understanding of the different mechanisms involved in growth and evolution of bacteria. I will specifically focus on the evolution of antibiotic resistance with an emphasis on the topics treated by the mathematical models used in this thesis. This chapter is thus not a general introduction to the microbiology of bacterial evolution, but rather a description of the concept of bacterial evolution from the perspective of a mathematician.

3.1 Growth of bacteria

To understand and model bacterial evolution it is necessary first to understand and be able to model bacterial growth.
Bacteria grow by dividing into two daughter cells. This process demands energy which comes from different substrates such as glucose. The growth of a bacterial population is typically characterized by four subsequent phases: the lag phase, the exponential growth phase, the stationary phase, and the death phase.
The mechanisms occurring during the **lag phase** are not fully understood, but it is believed that during this phase the bacteria adjust to the (new) environment, and enzymes, RNA, and proteins are synthesized to make the bacteria ready for cell division. The length of the lag phase depends on several factors, and includes the time for recovery from the physical damage or shock, which the bacteria experience when they are transferred to a new media. This phase also includes the time required for the synthesis of enzymes, which are necessary to catalyze the consumption of the new substrates.
During **exponential growth** the bacterial population grows by cell division. The rate of growth is specific for a given bacteria species and depends on the growth conditions. The growth can be parameterized by the growth rate ν or doubling time T_2 which are related by $T_2 = \log(2)/\nu$.
The **stationary phase** begins when either the substrate is depleted or when the

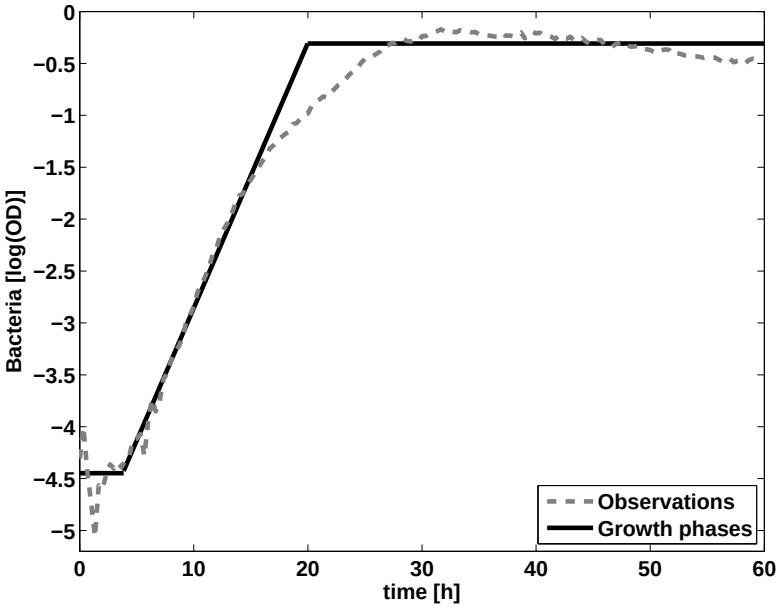


Figure 3.1: The logarithm of the size of a *Enterococcus* population as a function of time (dotted line) and a sketch of the three first phases of bacterial growth (full line): lag phase, exponential growth, and stationary phase.

maximum population size for the system has been reached. During the stationary phase the size of the population does not change anymore.
The final phase is the **death phase**, during which the population is reduced due to the death of bacteria. Only the three first phases are considered in this thesis, and experimental data after the onset of the death phase is therefore disregarded.

3.1.1 Modelling of growth

The picture drawn of bacterial growth in the previous section is simplified as shown in Figure 3.1, where a real growth curve for *Enterococcus* growing in Brain Heart Infusion (BHI) media is plotted together with an outline of the three first growth phases. Often the growth of a bacterial population during the exponential phase is not stationary, and different suggestions have been made to describe this non-stationary growth mathematically. The two most common types of models differ in the way they describe the onset of the stationary phase.
1) If the onset of the stationary phase occurs when the carrying capacity of the environment has been reached the logarithm of the bacterial concentration is typically modelled by a sigmoid curve (Baty and Delignette-Muller, 2004). The sigmoid curves are parameterized by a lag phase, λ, a maximum growth rate, $\mu_{\max}$, and a maximum bacterial concentration, $N_{\max}$. Several sigmoid curves have been suggested in literature (see Zwietering et al. (1990) and Baty and Delignette-Muller (2004) for comparisons of different models). Especially the modified Gompertz

curve has been broadly used, however it has been shown that the Baranyi model gives the best fit for a variety of datasets (Baty and Delignette-Muller, 2004). The Baranyi curve is given by

$$y(t) = y_{\max} + \log\left(\frac{-1+\exp(\mu_{\max}\lambda)+\exp(\mu_{\max}t)}{(-1+\exp(\mu_{\max}t))+\exp(\mu_{\max}\lambda+y_{\max}-y_0)}\right), \tag{3.1}$$

where $y = \log(N/N_0)$, $y_{\max} = \log(N_{\max})$ and N is the bacterial concentration.
2) A substrate dependent model should be considered when not enough substrate is available to reach intolerable numbers of bacteria before the growth rate decreases due to substrate depletion (Zwietering et al., 1990). Substrate dependent models are typically parameterized by a maximum growth rate $\mu_{\max}$, a yield factor or biomass yield coefficient $1/\eta$ and parameters defining the change of the growth rate as a function of substrate, S. This modelling framework does not include a description of the lag phase. The most widely used model applies the Monod equation for growth $(\mu_{\max}S)/(\kappa+S)$, where κ is the substrate concentration for which the growth is half of its maximum value. The model is often described by a set of coupled ODEs

$$\frac{dN}{dt} = \frac{\mu_{\max}S}{\kappa+S}N \tag{3.2}$$

$$\frac{dS}{dt} = -\eta\frac{\mu_{\max}S}{\kappa+S}N \tag{3.3}$$

where N is the bacterial concentration. In the experiments considered in this thesis it can be assumed that the growth enters stationary phase due to substrate depletion, and the above set of ODEs have therefore been the starting point of the model development carried out in Paper C. The Monod model is adequate for describing the growth on a single substrate, but it fails to model growth on rich media. In Paper C it is shown that growth on rich media can be better modelled with a growth proportional to the substrate level, i.e. υS, where υ is the specific growth rate and $S \in [0,1]$. Furthermore, it is shown that an extra state E for the enzyme level should be added to the model, to account for the lag-phase, resulting in the following set of SDEs

$$d\begin{bmatrix} B_t \\ S_t \\ E_t \end{bmatrix} = \begin{bmatrix} E_t\upsilon S_t B_t \\ -\eta E_t\upsilon S_t B_t \\ (\upsilon+\beta)S_t - E_t E_t \upsilon S_t - \beta E_t \end{bmatrix} dt + \begin{bmatrix} \sigma_B u_t & 0 & 0 \\ -\eta\sigma_B u_t & 0 & 0 \\ 0 & 0 & \sigma_E \end{bmatrix} d\omega\,. \tag{3.4}$$

where B_t is the bacteria concentration, u_t is an input of the measured bacteria concentration, and σ_B, σ_E, and σ are constants. The inclusion of the enzyme level in the model is inspired by the so-called optimal model or cybernetic model, which has been developed to describe growth on more than one substrate (Bajpai-Dikshit et al., 2003).

Fitness is a term often used in relation to bacterial growth and evolution. It is a measure of how fit a bacteria is, and thus how high the specific growth rate

indeed is. The fitness of a bacteria is highly dependent on the environment, where the growth takes place. For instance a resistant bacteria has a higher fitness than a sensitive bacteria in an environment with antibiotics, whereas the opposite might be the case if no antibiotics is present. In Paper B a Maximum Likelihood based method is described, which can be used to compare the growth rate for different bacteria in order to examine if they have a similar fitness. A bacteria can increase or decrease its fitness to an environment by different mechanisms as described in the next section.

3.2 Antibiotic resistance

Antimicrobial resistance is an increasing challenge resulting in difficulties for treatments of bacterial infections in both animals and humans. Sensitive bacteria will either have a low fitness or die in an environment with antibiotics. However, bacteria can acquire resistance by e.g. increasing the efflux pump activity or altering the target site for the antibiotics. Genes coding for mechanisms which result in antibiotic resistance can be transferred by either vertical or horizontal gene transfer. Gene transfer is said to be vertical when a cell divides and the genetic material is copied to the two daughter cells. By mutating a cell can acquire resistance and can then pass it on to next generation. Mutation is discussed in Section 3.2.1, while the process of horizontal gene transfer is discussed in Section 3.2.2 and 3.2.3.

3.2.1 Mutation and mutators

Spontaneous mutations occur in bacteria at rates of about $10^{-10} - 10^{-7}$ per cell devision (Drake, 1991; Boe et al., 2000). The development of resistance requires one or several mutations, which according to classical Darwinism each has to be established in the population before new mutations can arise. This is a very slow process, which by itself can not explain the actual evolution of bacteria. However, bacteria can increase the mutation frequency for example by a mutation in the genes responsible for DNA repair (the mismatch repair genes). Bacteria with such an increased mutation rate are called **mutators**, and they are believed to play an important role in the development of antibiotic resistance. Mutator bacteria are found with increasing frequencies in natural and clinical environments as for instance in cystic fibrosis patient. This has motivated the study in Paper E of mutators in a *Pseudomonas aeruginosa* population. A competition experiment between a mutator bacteria and the corresponding wild-type bacteria (bacteria with normal mutation rate) has been performed, and an SDE based model is developed to describe the dynamics of the system. From experiments it is seen that the *mutY-mutM* mutant will take over the bacterial population when in competition with the wild-type bacteria strain. The model suggests that this is caused by a higher rate of adaption in the mutator. Furthermore the model predicts that for the other mutators considered in the experiment (*mutM*, *mutT*, and *mutY*) a lower equilibrium mutator frequency will be obtained, as the equilibrium frequency is depen-

dent on the mutator strength. The study suggests that the high ratio of mutators found in cystic fibrosis patients can be caused be a low concentration of antibiotics in the lungs, to which mutators can adapt faster.
In Paper F it is shown, how mutator populations can function as "genetic work stations", where multiple mutations occur and subsequently are transmitted to the parent population by horizontal gene transfer. Thus, a mutator population can be seen as a way to bypass traditional Darwinian evolution. This new hypothesis for the role of mutator bacteria in the evolution of resistance is supported by our stochastic model.

3.2.2 Horizontal gene transfer

There are three different mechanisms of horizontal gene transfer: transduction, transformation, and conjugation. By **transduction** genetic material is passed on from one cell to another via a bacteriophage. The phage enters the first host cell where it during replication might take up some of the cells own DNA. At some point the bacteriophage will leave the cell ready for infecting a new host cell. When entering a new cell the DNA from the first cell located in the bacteriophage might be incorporated in the chromosome of the new cell during recombination. The second mechanism, **transformation**, is the uptake of free DNA from the environment, such as DNA released by a dead cell. The third mechanism, **conjugation**, is the transfer of small DNA molecules called plasmids from a donor to a recipient by cell-to-cell contact. The donor and recipient cells establish a mating pair during which the plasmid is copied to the recipient, which thereby turns into a transconjugant. Conjugation is the most effective method for the transfer of antibiotic resistance provided that the resistance is located on plasmids. Therefore conjugation is the main focus of the horizontal gene transfer considered in this thesis and in the next section.

3.2.3 Modelling of conjugation

Conjugation can be modelled in different ways depending on the purpose of the model and the degree of information available. Several authors (Levin and Stewart, 1980; Freter et al., 1983; Knudsen et al., 1988; Clewlow et al., 1990; Top et al., 1992) model conjugation with the mass action model proposed by Levin et al. (1979), which states that plasmid transfer is proportional to the concentrations of donor and recipient, i.e. γDR, where γ is the conjugation rate, D is the concentration of donor bacteria, and R is the concentration of recipient bacteria. In Paper F the mass action model is used to describe conjugation for a simulation study. However, the concentration of the recipient population is assumed to be constant for the model, which leads to a fixed conjugation rate per donor bacteria. This is in agreements with studies by Andrup et al. (1998), who has shown that the conjugation rate per donor reaches a maximum level for recipient concentration above 10^7 cells/ml.

Paper D presents an alternative description for the mass action model for growth in rich exhaustible liquid media,

$$\frac{\gamma_{\max} S}{\kappa_c + S} DR\,, \tag{3.5}$$

For *Enterococcus faecium* which was examined in Paper D, the conjugation rate is found to be maximum until the substrate is totally depleted, whereafter the rate falls to zero. The κ_c value for conjugation is found to be much lower than the κ value in the Monod growth term used to model growth. The model therefore suggests that conjugation continues after the stationary phase has been reached.

3.3 Experimental methods

In the laboratory bacteria can grow in liquid media or on agar plates (petri dishes containing growth media in a gel). When bacteria grow in liquid media there are different ways of measuring the number of bacteria. An estimate of the bacterial concentration can be made by taking a sample of a known volume and distributing that over an agar plate. Each bacteria in the sample will grow from the nutrition on the plate and form a so-called colony forming unit (CFU) which can be seen by eye after approximately 12 hours. Each colony corresponds to one bacterium in the original sample. Up to 200 colonies can be separated from each other on the plate, which means that the sample typically will have to be diluted in order to reduce the number of colonies. To obtain a good estimate of the concentration in the liquid the analysis is often made in duplicate for several dilutions. The concentration in the liquid can then be determined with Poisson regression as described in Paper D. The agar plate can also be made selective for a certain bacteria strain by adding for instance antibiotics to the plate. Bacteria can be constructed for an experiment with different resistance patterns, such that they can be separated by counting on different selective plates. This technique has been used for the experiments described in Paper D and Paper E. While performing the experiments for the study in Paper D we found that the donor and recipient could conjugate for some time on the plates selecting for transconjugants, i.e. agar plates with antibiotics which should kill the donor and recipient bacteria. This methodological problem was solved by including conjugation on the plate in the measurement equation of the model, whereby the model was used to separate conjugation in the liquid from conjugation on the plate. To our knowledge we are the first to recognize this problem (even though it must occur in many experiments), and a solution is suggested, such that a correct estimate can be made for the conjugation rate.

The disadvantage of estimating concentrations by the CFU count method is the large amount of work required for taking the samples, performing dilution and plating. Therefore turbidimetric instruments such as bioscreens are often used to measure the optical density (OD) of the solution and the measurements can subsequently be transferred into an estimate of the bacteria concentration. Paper A

contains a comparison of different calibration curves for the relation between the OD and CFU values, and the study concludes that the exponential curve

$$\mathrm{OD} = a \cdot (1 - \exp(b \cdot \mathrm{CFU})) , \tag{3.6}$$

best describes the relation. This is a very strong relation , as it is derived from physical principles and can be extended for higher cell concentrations.
Advantages of bioscreen measurements are the fast sampling time and the possibility of running several experiments simultaneously. Therefore this method has been used for most of the experimental studies considered in this thesis.

3.4 Spread of antibiotic resistant bacteria

The evolution of antibiotic resistance is followed by an increasing prevalence of resistance in human populations (and in animals, but this will not be considered here). In Figure 3.2 the prevalence of *E. coli* resistant to 3rd generation cephalosporins is shown for four European countries. Resistance to 3rd generation cephalosporins can be seen as a representation of extended-spectrum beta-lactamases (ESBL) resistance. The prevalence is seen to be increasing, which gives a challenge for the treatment of infections. In the latest DANMAP rapport (from 2008) from the Danish integrated antimicrobial resistance monitoring and research programme one of the focus areas was the increasing prevalence of ESBL producing *Klebsiella pneumonia* in Denmark (www.danmap.org). The report concludes that studies are needed to improve the understanding of the spread of resistant bacteria.

The spread of ESBL producing *E. coli* has been the attention of the study in Paper G. A model has been developed to describe the spread of ESBL producing *E. coli* in a hospital and its catchment population. The model includes external inflow of resistance with travellers, cross-transfer of resistance, mutation and conjugation. The dynamics is considered only on the level of human individuals, and do not contain any information about the number of bacteria of a certain type in each person. This means that the more advanced models for mutation and conjugation developed in this thesis can not be transferred directly to this study. Conjugation, for instance, is therefore implemented as a probability of occurrence in the human intestine including the subsequent growth of the bacteria to reach detectable amounts. This parametrization has been chosen due to a lack of data which render a more adequate model impossible. The ESBL model enables a study of the importance of different transmission routes, and can hereby help in the ongoing process of diminishing the spread of resistance.

3.5 Discussion

The cross-disciplinary work between a mathematician/statistician and microbiologist/medical doctor has the advantage that it may provide insights beyond the directly measurable scientific outcome of mathematical and statistical analysis. A

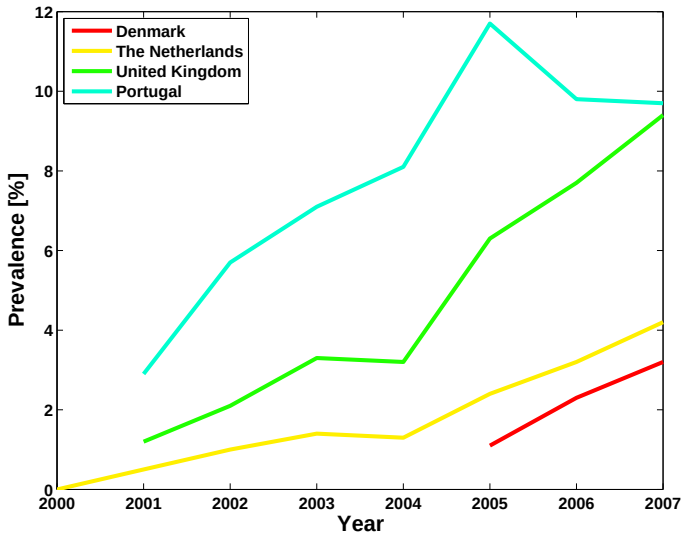


Figure 3.2: The prevalence of *E. coli* resistant to 3rd generation cephalosporins, which can be seen as a representation of ESBL producing bacteria.

mathematical modeller will inevitably ask a lot of questions to be able to make a good parametrization of the biological system; in the case of this thesis: bacterial evolution and spread. These questions will lead the microbiologist to consider new explanations and tests for the dynamics of the system. The modeller on the other hand will get new ideas for the model development, which can be tested using various statistical tools as discussed in the previous chapter. Due to benefits of this symbiotic relation it is advisable to combine experimental studies with mathematical modelling whenever possible.

Based on the result from the conjugation model in Paper D it is advised to examine for conjugation on the selective plates whenever performing a conjugation experiment, simply by mixing the donor and recipient strains directly on the plates. If conjugation is found to occur on the plates the extended model proposed here should be used.

The competition model developed as part of this thesis gives a good representation of the *in vitro* competitive growth between mutator and wild-type *P. aeruginosa*. It could be very interesting to apply the competition model on data from *in vivo* studies or to adjust the mathematical model to the environments in human lungs under treatment with ciprofloxacin.

For two studies described in this thesis there have been no or very sparse data. This is the case with the simulation study for the transfer via a mutator subpopulation (Paper F). It could be interesting to get the simulation results confirmed

by experiments. The other study is the model for the transfer of ESBL (Paper G), where the parameters of a model with 20 states are based on only 9 data points. It is therefore recommended to perform a sensitivity analysis for the different model parameters. Further simulation studies with the ESBL model could be used to examine the consequence of different interventions in the hospital, but it is advised to confirm the reliability of the model with more extensive data, before continuing with intervention simulation studies.

CHAPTER

4 Conclusion

During this project it has been shown that mathematical modelling and statistics are very useful tools for understanding the evolution of antibiotic resistance and the spread of resistance. The use of SDEs as a grey-box modelling framework to study bacterial populations is a new approach, which has given promising results during this project.
SDEs can, opposed to traditionally used models based on ODEs, describe autocorrelated residuals in the data. Not describing the autocorrelation structure by the model could lead to false conclusions of statistical tests, as the model is not correctly parameterized. However, using SDEs opens up for the use of strong statistical tools. Hence, in Paper C a new SDE framework for systematic model development is made and applied to suggest a new expression for growth in rich media.

An SDE based state-space model has been developed in Paper D and it is successfully applied to model conjugation in a broth exhaustible media. The maximum likelihood based framework for estimating model parameters combined with likelihood-ratio tests and Akaike's information criterion for inference studies used in this paper provides strong tools for model improvements. In the same paper a new expression for substrate dependent conjugation is described which can be used to model conjugation in a broth media. During the process of developing a suitable model for conjugation a methodological problem was encountered: the fact that conjugation occur not only in the broth but also on the plates selecting for transconjugants. This was solved using the grey-box modelling framework by including plate conjugation in the observation equation. Hereby the plate conjugation can be separated from conjugation in the broth media. Even though this extension to the model did not perform significantly better for the given data, it is believed that the method will be the preferred choice if more frequent sampled data is available for the analysis.

A competition experiment has been successfully described with a new SDE based model described in Paper E, from which it can be inferred that mutator bacteria will have an advantage in an environment with low ciprofloxacin concentrations, due to their higher mutation rate and thereby ability to adjust to the environment. The study uses OD measurements, which are converted to cell concentrations. The conversion is made using an exponential function based on physical principles, which has been found to give the best result compared to other functions (Paper A). The reliability of the competition study is dependent on the mutator and non-mutator strains to have the same specific growth rates. This was indeed proven to be the case when the linear regression models for the bacterial concentration containing an exponential decaying autocorrelation function and state dependent variance was compared using a likelihood-ratio test (Paper B). Finally the competition model was based on an earlier study (Paper C) for the implementation of growth in rich media. The growth rate for bacteria growing on minimal media is traditionally modelled using the Monod equation, but in Paper C it is found that a growth rate proportional to the substrate level gives a better fit to the data.

The model for the transfer of bacterial resistance via a mutator subpopulation described in Paper F is an excellent example of the strength of mathematical modelling. In the paper it is examined whether mutator bacteria can function as "genetic work stations", where multiple mutations occur and are subsequently transmitted to the non-mutator population by horizontal gene transfer. It is found that mutator populations can be seen as a way to bypass traditional Darwinian evolution. This theory is intriguing, as it can help explain the speed by which evolution occurs, and at the same time give a good explanation for the high ratio of mutator bacteria found in natural environments. It has not been possible to demonstrate the theory by experimental studies, but with our stochastic model we are able to provide indications that this route of transfer actually does take place.

The development of antibiotic resistance provides an increasing challenge for treatments of bacterial infections. Before being able to slow down this process, it is necessary to understand the mechanisms leading to resistance. This thesis has contributed to a better understanding of the evolution process, especially the role of mutator bacteria.

A patient with a bacterial infection can in many cases obtain resistance not only by evolution of bacteria in the body, but also from external sources, such as other patients. In Paper G a model is proposed which shows the importance of different routes for acquisition of ESBL resistance in *E. coli*. The model concludes that cross-transfer plays a significant role in the increasing prevalence of resistant bacteria. Thus, a good hygiene at, e.g., hospitals is a key component for reducing the prevalence of resistant bacteria.

CHAPTER

5 Bibliography

5.1 Bibliography

Andersen, K. K., Madsen, H., Hansen, L. H., 2000. Modelling the heat dynamics of a building using stochastic differential equations. Energ. Buildings 31 (1), 13–24.

Andrup, L., Andersen, K., 1999. A comparison of the kinetics of plasmid transfer in the conjugation systems encoded by the F plasmid from Escherichia coli and plasmid pCF10 from Enterococcus faecalis. Microbiology 145 (8), 2001–2009.

Andrup, L., Smidt, L., Andersen, K., Boe, L., 1998. Kinetics of conjugative transfer: A study of the plasmid pxo16 from *Bacillus thuringiensis* subsp. *israelensis*. Plasmid 40 (1), 30–43.

Augustin, J.-C., Rosso, L., Carlier, V., 1999. Estimation of temperature dependent growth rate and lag time of listeria monocytogenes by optical density measurements. J. Microbiol. Methods 38 (1-2), 137–146.

Bajpai-Dikshit, J., Suresh, A. K., Venkatesh, K. V., 2003. An optimal model for representing the kinetics of growth and product formation by lactobacillus rhamnosus on multiple substrates. J. Biosci. Bioeng. 96 (5), 481–486.

Baty, F., Delignette-Muller, M.-L., 2004. Estimating the bacterial lag time: which model, which precision? Int. J. Food Microbiol. 91, 261–277.

Baty, F., Flandrois, J. P., Delignette-Muller, M. L., 2002. Modeling the lag time of *Listeria monocytogenes* from viable count enumeration and optical density data. Appl. Environ. Microbiol. 68, 5816–5825.

Boe, L., Danielsen, M., Knudsen, S., Petersen, J. B., Maymann, J., Jensen, P. R., 2000. The frequency of mutators in populations of *Escherichia coli*. Mutat. Res. 448 (1), 47–55.

Chorin, E., Thuault, D., Cléret, J.-J., Bourgeois, C.-M., 1997. Modelling bacillus cereus growth. Int. J. Food Microbiol. 38 (2-3), 229–234.

Christiansen, L. E., 2004. Nonlinear time series analysis of zoonoses. Ph.D. thesis, Department of Informatics and Mathematical modelling, Technical Univisity of Denmark.

Clewlow, L. J., Cresswell, N., Wellington, E. M. H., 1990. Mathematical Model of Plasmid Transfer between Strains of Streptomycetes in Soil Microcosms. Appl. Environ. Microbiol. 56 (10), 3139–3145.

Dalgaard, P., Koutsoumanis, K., 2001. Comparison of maximum specific growth rates and lag times estimated from absorbance and viable count data by different mathematical models. J. Microbiol. Methods 43 (3), 183–196.

Drake, J. W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U. S. A. 88 (16), 2357641–7164.

Freter, R., Freter, R. R., Brickner, H., 1983. Experimental and Mathematical Models of Escherichia coli Plasmid Transfer In Vitro and In Vivo. Infect. Immun. 39 (1), 60–84.

Frydendall, J., 2009. Particle filters with applications. Tech. rep., Technical University of Denmark, Department of Informatics and Mathematical Modeling, Mathematical Statistics.

Fujikawa, H., Kai, A., Morozumi, S., 2004. A new logistic model for *Escherichia coli* growth at constant and dynamic temperatures. Food Microbiol. 21 (5), 501–509.

Gelb, A., Kasper Jr., J. F., Nash Jr., R. A., Sutherland Jr., A. A., 1974. Applied Optimal Estimation. The M.I.T. press.

Ionides, E. L., Bretó, C., King, A. A., 2006. Inference for nonlinear dynamical systems. Proc. Natl. Acad. Sci. U. S. A. 103 (49), 18438–18443.

Jacobsen, J. L., Madsen, H., Harremoes, P., 1997. A stochastic model for two-station hydraulics exhibiting transient impact. Water Sci. Technol. 36 (5), 19–26.

Jimenez, M. J., Madsen, H., Bloem, J. J., Dammann, B., 2008. Estimation of nonlinear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules. Energ. Buildings 40 (2), 157–167.

Jonsdottir, H., Jacobsen, J. L., Madsen, H., JUN 2001. A grey-box model describing the hydraulics in a creek. Environmetrics 12 (4), 347–356.

Juneja, V., Valenzuela Melendres, M., Huang, L., Gumudavelli, V., Subbiah, J., Thippareddi, H., 2007. Modeling the effect of temperature on growth of *Salmonella* in chicken. Food Microbiol. 24 (4), 328–335.

Klim, S., Mortensen, S. B., Kristensen, N. R., Overgaard, R. V., Madsen, H., 2009. Population stochastic modelling (psm)-an r package for mixed-effects models based on stochastic differential equations. Comput. Methods Programs Biomed. 94 (3), 279–289.

Knudsen, G. R., Walter, M. V., Porteous, L. A., Prince, V. J., Armstrong, J. L., Seidler, R. J., 1988. Predictive model of conjugative plasmid transfer in the rhizosphere and phyllosphere. Appl. Environ. Microbiol. 54 (2), 343–347.

Kovárová-Kovar, K., Egli, T., 1998. Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics. Microbiol. Mol. Biol. Rev. 62 (3), 646–666.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004a. A method for systematic improvement of stochastic grey-box models. Comput. Chem. Eng. 28, 1431–1449.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004b. Parameter estimation in stochastic grey-box models. Automatica 40 (40), 225–237.

Øksendal, B., 2007. Stochastic Differential Equations. Springer.

Law, A. M., Kelton, W. D., 2000. Simulation modeling and analysis, third edition Edition. McGraw-Hill.

Lele, S. R., Dennis, B., Lutscher, F., 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods. Ecol. Lett. 10 (7), 551–563.

Levin, B., F.M., S., Rice, V., 1979. The kinetics of conjugative plasmid transmission: fit of a simple mass action model. Plasmid 2 (2), 247–260.

Levin, B. R., Stewart, F. M., 1980. The population biology of bacterial plasmids: A priori conditions for the existence of mobilizable nonconjugative factors. Genetics 94 (2), 425–443.

Li, H., Xie, G., Edmondson, A., 2007. Evolution and limitations of primary mathematical models in predictive microbiology. Brit. Food J. 109 (8), 608–626.

Lindqvist, R., 2006. Estimation of *Staphylococcus aureus* growth parameters from turbidity data: Characterization of strain variation and comparison of methods. Appl. Environ. Microbiol. 72 (7), 4862–4870.

Liu, Y., 2006. A simple thermodynamic approach for derivation of a gereral monod equation for microbial growth. Biochem. Eng. J. 31, 102–105.

Lobry, J., Flandrois, J., Carret, G., Pave, A., 1992. Monod's bacterial growth model revisited. B. Math. Biol. 54, 117–122.

López, S., Prieto, M., Dijkstra, J., Dhanoa, M., France, J., 2004. Statistical evaluation of mathematical models for microbial growth. Int. J. Food Microbiol. 96 (3), 289–300.

Ma, W. T., Sandri, G. V., Sarkar, S., 1992. Analysis of the luria-delbrück distribution using discrete convolution powers. J. Appl. Prob. 29 (2), 255–267.

MacDonald, Jacqueline, A., Smets, Barth, F., Rittmann, Bruce, E., 1992. Effects of energy availability on the conjugative-transfer kinetics of plasmid rp4. Water Res. 26 (4), 461–468.

Madsen, H., 2008. Time series analysis. Chapman & Hall/CRC.

Madsen, H., Holst, J., 1995. Estimation of continuous-time models for the heat dynamics of a building. Energ. Buildings 22 (1), 67–79.

Madsen, H., Thyregod, P., 1988. Modelling the time correlation in hourly observations of direct radiation in clear skies. Energ. Buildings 11, 201–211.

Mandsberg, L. F., Ciufo, O., Christensen, L. E., Højby, N., 2008. Antibiotic resistance in hypermutable *P. aeruginosa* due to inactivation of the dna oxidative repair system, submitted to: Antimicrobial Agents and Chemotherapy.

Monod, J., 1949. The growth of bacterial cultures. Annu. Rev. Microbiol. 3, 371–394.

Montanari, S., Oliver, A., Salerno, P., Mena, A., Bertoni, G., Tummler, B., Cariani, L., Conese, M., Doring, G., Bragonzi, A., 2007. Biological cost of hypermutation in *Pseudomonas aeruginosa* strains from patients with cystic fibrosis. Microbiology - Reading 153 (5), 1445.

Mortensen, S. B., Klim, S., Dammann, B., Kristensen, N. R., Madsen, H., Overgaard, R. V., 2007. A matlab framework for estimation of nlme models using stochastic differential equations. J. Pharmacokinet. Pharmacodyn. 34 (5), 623–642.

Nielsen, J. N., Vestergaard, M., Madsen, H., 2000. Estimation in continuous-time stochastic volatility models using nonlinear filters. International Journal of Theoretical and Applied Finance 3 (4), 279–308.

Novais, A., Canton, R., Coque, T. M., Moya, A., Baquero, F., Galan, J. C., 2008. Mutational events in cefotaximase extended-spectrum beta-lactamases of the ctx-m-1 cluster involved in ceftazidime resistance. Antimicrob. Agents Chemother. 52 (7), 2377–2382.

Pedersen, M. W., Righton, D., Thygesen, U. H., 2008. Geolocation of north sea cod (gadus morhua) using hidden markov models and behavioural switching. Can. J. Fish. Aquat. Sci. 65 (11), 2367–2377.

Quenee, L., Lamotte, D., Polack, B., 2005. Combined sacb-based negative selection and cre-lox antibiotic marker recycling for efficient gene deletion in pseudomonas aeruginosa. BioTechniques 38 (1), 63–67.

Shama, G., Perni, S., Andrew, P., 2005. Estimating the maximum growth rate from microbial growth curves: definition is everything. Food Microbiol. 22 (6), 491–5.

Sudarshana, P., Knudsen, G. R., 2006. Quantification and modeling of plasmid mobilization on seeds and roots. Curr. Microbiol. 52 (6), 455–459.

Top, E., Vanrolleghem, P., Mergeay, M., Verstraete, W., 1992. Determination of the mechanism of retrotransfer by mechanistic mathematical modeling. J. Bacteriol. 174 (18), 5953–5960.

Willms, A., Roughan, P., Heinemann, J., 2006. Static recipient cells as reservoirs of antibiotic resistance during antibiotic therapy. Theor. Popul. Biol. 70 (4), 436–451.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., Riet, K. v., 1990. Modeling of the bacterial growth curve. Appl. Environ. Microbiol. 56 (6), 1875–1881.

Appendices

PAPER

A

Comparison of calibration curves for the relation between optical density and viable cell count data [‡]

Abstract

Optical density is often used as a measurement of bacterial concentration. In this study five calibration curves relating optical density measurements to colony-forming unit measurements of bacterial concentration are compared for *P. aeruginosa* and *E. coli*. Using Maximum Likelihood estimation and Akaike's information criterion to compare the curves it is found that an exponential function derived from physical principles is the best calibration curve.

In many areas of microbiology there is a need for large quantities of accurate and high-quality data. Therefore turbidimetric instruments, e.g. bioscreens, are often used to measure the optical density (OD) of bacteria and then provide an estimate for the bacterial density.

In the literature, several calibration curves have been suggested for converting OD measurement into viable count or colony-forming units (CFU). Correct conversion is very important when using OD data for estimating characteristic growth parameters such as the maximum growth rate, biomass yield coefficient or maximum bacterial concentration, or when using the OD data to develop new mathematical models. In this study we compare the following five calibration curves suggested in earlier studies for *Pseudomonas aeruginosa* and *Escherichia coli*:

Linear (Corman et al., 1986; Augustin et al., 1999; Baty et al., 2002):

$$\mathrm{OD}_i = a \cdot \mathrm{CFU}_i + \epsilon_i \tag{A.1}$$

[‡]Submitted as: K. R. Philipsen, L. E. Christiansen, L. F. Mandsberg, H. Hasman, H. Madsen, 2010. Comparison of calibration curves for the relation between optical density and viable cell count data. Submitted to *Applied and Environmental Microbiology*.

Second-order (McClure et al., 1993):

$$\mathrm{OD}_i = a \cdot \mathrm{CFU}_i + b \cdot \mathrm{CFU}_i^2 + \epsilon_i \tag{A.2}$$

Third-order (Stephens et al., 1997):

$$\mathrm{OD}_i = a \cdot \mathrm{CFU}_i + b \cdot \mathrm{CFU}_i^2 + c \cdot \mathrm{CFU}_i^3 + \epsilon_i \tag{A.3}$$

Logarithmic (Chorin et al., 1997; Valero et al., 2006; Francois et al., 2005):

$$\mathrm{OD}_i = \exp\left(\frac{\log(\mathrm{CFU}_i) - a}{b}\right) + \epsilon_i \tag{A.4}$$

Exponential (Christiansen, 2004):

$$\mathrm{OD}_i = a \cdot (1 - \exp(-b \cdot \mathrm{CFU}_i)) + \epsilon_i \tag{A.5}$$

Letting N be the size of each dataset, then $i = 1, \ldots, N$, and the residual $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_N]$ $\sim \mathcal{N}(0, \sigma^2\boldsymbol{\Sigma})$ for all models. $\sigma^2\boldsymbol{\Sigma}$ is the covariance matrix of the N OD measurements.

The direct linear relationship between OD and CFU data is supported by Beer's law, which states that the amount of absorbed radiation is proportional to the concentration in the sample. The exponential calibration curve (A.5), which was first introduced by Christiansen (2004), accounts for the shadow effect caused by bacteria at high OD values. This shadow effect results in an underestimation of the bacterial concentration. The theoretical motivation for this new expression is that an OD measurement is made by transmitting light through a sample so that the OD measurement is actually made by a projection of the sample onto a plane. If a given proportion, p, of the cross section is covered by bacteria and one additional bacteria is added, then on average p times the cross section of the bacteria, a, is already covered by other bacteria and thus only $(1-p)a$ is added to the overall coverage. This leads to Equation (A.5). It should be noted that when p is small, i.e. for low cell concentrations, the relationship between OD and CFU is approximately linear.

Experiments were made in order to obtain data to compare the calibration curves. *P. aeruginosa* (PAO1 from Stover et al. (2000)) was taken from an agar plate and diluted in salt water ($NaCl_2$ 0.9%) until an OD600 $\approx$ 1.8 was reached. *E.coli* (MG1655) was grown overnight whereafter the sample was spinned down and diluted to obtain an OD600 $\approx$ 1.8. This last method enabled to reach a higher concentration. Hereafter appropriate dilutions were made. The OD of each dilution was measured in the bioscreen (Labsystem C (Bie og Berntsen)) for two hour and the CFU of each dilution was counted on Luria-Bertani plates. From the CFU counts, the expected concentration in the sample was calculated by Poisson regression, as explained by Philipsen et al. (2010). From the OD measurements, the OD of a well containing only saltwater was subtracted and the mean over the two hour measurements was found.

We use a Maximum Likelihood (ML) approach (Philipsen et al., 2008; Madsen, 2008) to estimate the curve parameters and compare the curves. Assuming that the data is normally distributed (this should be checked by examining the distribution of the residuals) the log-likelihood function to be optimized is

$$\log(\mathscr{L}(\boldsymbol{\theta}|\boldsymbol{Y})) = -\frac{1}{2}N\log(\sigma^2) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2\sigma^2}(\mathbf{OD}-\widehat{\mathbf{OD}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{OD}-\widehat{\mathbf{OD}})\,. \quad \text{(A.6)}$$

Here $\widehat{\mathbf{OD}}$ is a vector of estimated OD values from a calibration curve, and $\mathbf{OD}$ is a vector of the corresponding OD measurements. The residual variance σ^2 is estimated simultaneously with the parameter estimation. The studies considering the models, Equations (A.1)-(A.3), assume that the variance is the same for all OD measurements, i.e. $\boldsymbol{\Sigma} = \boldsymbol{I}$, and therefore the problem reduces to (unweighted) least squares. The logarithmic transformation Equation (A.4), with a linear relation between the log(OD) and the log(CFU), was originally introduced to normalize the variance. Instead of transforming the data a concentration-dependent variance can be included in the covariance matrix to weight the observations. The variance of the OD measurements increases with increasing OD values, which can be implemented in the covariance matrix as $\boldsymbol{\Sigma} = \text{diag}(\widehat{\mathbf{OD}})$. Both implementations of the covariance matrix will be tested. The optimum of the log-likelihood function is computed in `R` using the command `ucminf` (Nielsen and Mortensen, 2009).

The result of the parameter estimation for each calibration curve can be seen in

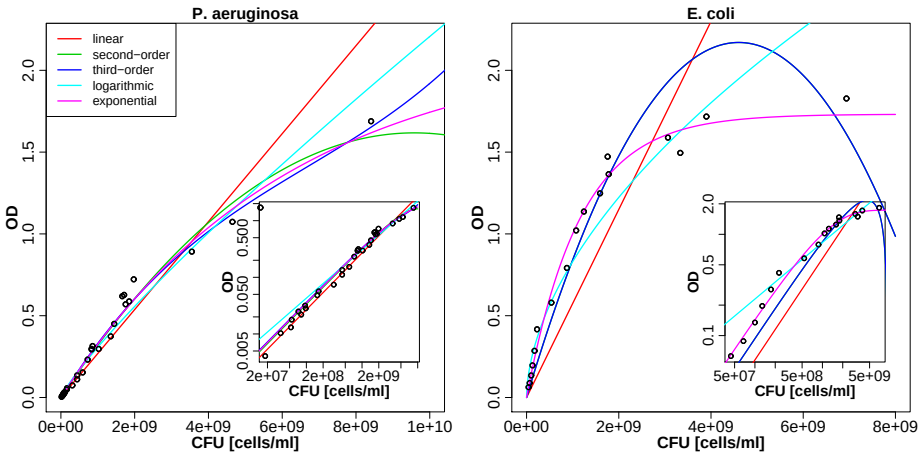
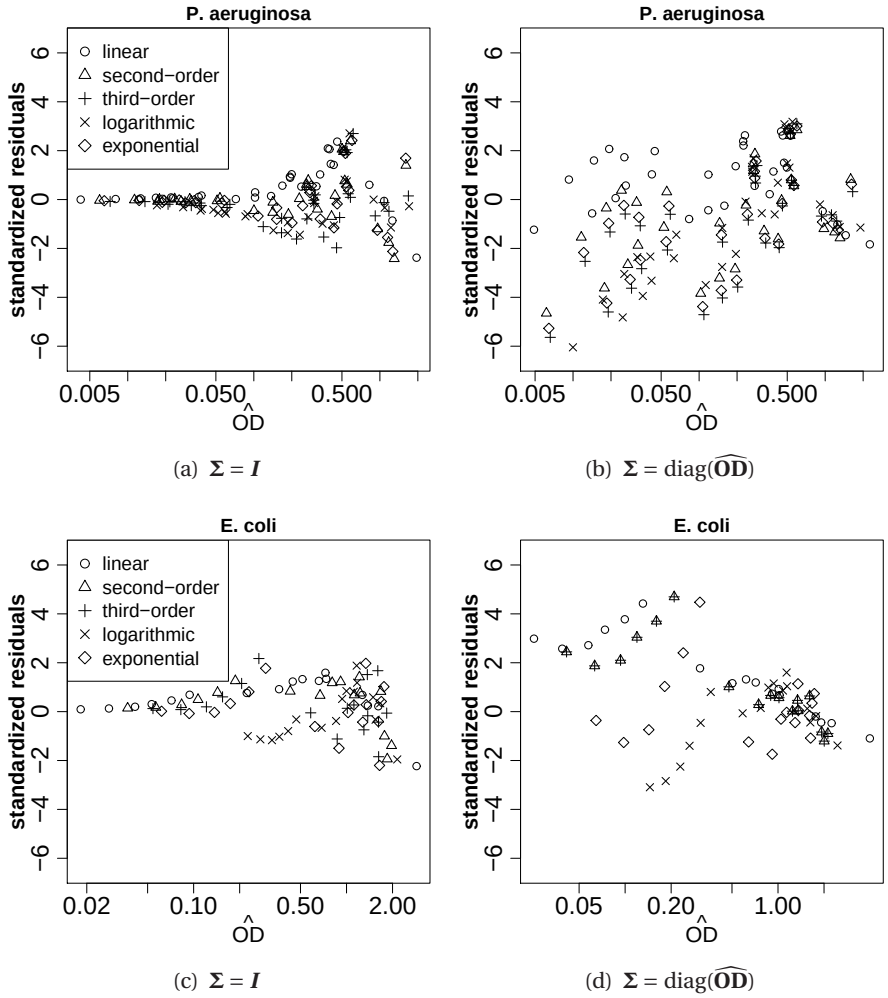


Figure A.1: Estimates of the calibration curves for $\boldsymbol{\Sigma} = \text{diag}(\widehat{\mathbf{OD}})$.

Figure C.7. An inference study for the different curves is performed using Akaike's Information Criterion (AIC). AIC is chosen, as it can be used to compare non-nested models in apposed to other methods such as the F-test (Zwietering et al., 1990; López et al., 2004) and the likelihood-ratio test (Philipsen et al., 2008, 2010). For all except one function, the AIC value when using a weighted estimation is



(a) $\boldsymbol{\Sigma} = \boldsymbol{I}$

(b) $\boldsymbol{\Sigma} = \text{diag}(\widehat{\mathbf{OD}})$

(c) $\boldsymbol{\Sigma} = \boldsymbol{I}$

(d) $\boldsymbol{\Sigma} = \text{diag}(\widehat{\mathbf{OD}})$

Figure A.2: The standardized residuals for each of the two correlation matrices and for each of the five calibration curves.

lower than for an unweighted estimation, which indicates that the weighted estimation should be used. This can be further confirmed by considering the standardized residuals as a function of the estimated OD values in Figure A.2. It is seen that the variance is indeed proportional to the estimated OD values (Figure A.2 left column). Hence, the correlation matrix $\boldsymbol{\Sigma} = \text{diag}(\widehat{\mathbf{OD}})$ gives the best description of the variance in the system (Figure A.2 right column).

As seen from Table A.1, the calibration curve with the lowest AIC value for $\boldsymbol{\Sigma} = \text{diag}(\widehat{\mathbf{OD}})$ is the exponential function. For *P. aeruginosa* the AIC values for the polynomials are close to the AIC for the exponential function, and all three functions thus describe the dataset well. For *E. coli* the exponential function describes the data better than the other tested functions. The polynomials can not

Table A.1: AIC for each of the five calibration curves tested.

	P. aeruginosa		*E. coli*	
$\boldsymbol{\Sigma}$	$\boldsymbol{I}$	diag($\widehat{\mathbf{OD}}$)	$\boldsymbol{I}$	diag($\widehat{\mathbf{OD}}$)
linear	-99.78	-146.91	-7.81	-20.98
second-order	-137.64	-178.78	-36.94	-43.46
third-order	-156.73	-179.87	-67.86	-41.49
logarithmic	-140.38	-158.39	-40.65	-43.50
exponential	-145.49	-180.65	-70.35	-74.30

be used for extrapolation to higher cell concentrations, whereas this is possible with the exponential function. Thus, on the basis of physical principles, the exponential curve is preferred. The estimated parameters for the exponential curve are for *P. aeruginosa:* $a = 2.17$, $b = 1.63 \cdot 10^{-10}$, $\sigma^2 = 0.005$; and for *E. coli:* $a = 1.73$, $b = 8.62 \cdot 10^{-10}$, $\sigma^2 = 0.078$. OD measurements can be translated to CFU by isolating CFU in Equation (A.5) and inserting the values for the constants a and b.
On the basis of this study, we recommend that the exponential function is applied to convert OD measurements into CFU data. The exponential calibration curve is based on physical principles and is found to give the best fit to the data.

This study was supported by the Danish Research Council for Technology and Production Sciences through Grant 274-05-0117.

A.1 Bibliography

Augustin, J.-C., Rosso, L., Carlier, V., 1999. Estimation of temperature dependent growth rate and lag time of Listeria monocytogenes by optical density measurements. J. Microbiol. Methods 38 (1-2), 137–146.

Baty, F., Flandrois, J., Delignette-Muller, M., 2002. Modeling the lag time of Listeria monocytogenes from viable count enumeration and optical density data. Appl. Environ. Microbiol. 68 (12), 5816–5825.

Chorin, E., Thuault, D., Cléret, J.-J., Bourgeois, C.-M., 1997. Modelling Bacillus cereus growth. Int. J. Food Microbiol. 38 (2-3), 229–234.

Christiansen, L. E., 2004. Nonlinear time series analysis of zoonoses. Ph.D. thesis, Department of Informatics and Mathematical modelling, Technical Univisity of Denmark.

Corman, A., Carret, G., Pave, A., Falndrois, J. P., Couix, C., 1986. Bacterial growth measurement using an automated system mathematical modeling and analysis of growth kinetics. Ann. Inst. Pasteur Mic. 137B (2), 133–144.

Francois, K., Devlieghere, F., Standaert, A., Geeraerd, A., Cools, I., Van Impe, J., Debevere, J., 2005. Environmental factors influencing the relationship between optical density and cell count for Listeria monocytogenes. J. Appl. Microbiol. 99 (6), 1503–1515.

López, S., Prieto, M., Dijkstra, J., Dhanoa, M., France, J., 2004. Statistical evaluation of mathematical models for microbial growth. Int. J. Food Microbiol. 96 (3), 289–300.

Madsen, H., 2008. Time series analysis. Chapman & Hall/CRC.

McClure, P. J., Cole, M. B., Davies, K. W., Anderson, W. A., 1993. The use of automated turbidimetric data for the construction of kinetic models. J. Ind. Microbiol. 12 (3-5), 277–285.

Nielsen, H. B., Mortensen, S. B., 2009. ucminf: General-purpose unconstrained non-linear optimization. R package version 1.0-5.

Philipsen, K., Christiansen, L., Hasman, H., Madsen, H., 2010. Modelling conjugation with stochastic differential equations. Journal of Theoretical Biology 263 (1), 134–142.

Philipsen, K. R., Christiansen, L. E., Mandsberg, L. F., Madsen, H., 2008. Maximum likelihood estimation of the specific groth rate for P. aeruginosa and four mutator strains. J. Microbiol. Methods 75 (3), 551–557.

Stephens, P., Joynson, J., Davies, K., Holbrook, R., Lappin-Scott, H., Humphrey, T., 1997. The use of an automated growth analyser to measure recovery times of single heat-injured Salmonella cells. J. Appl. Microbiol. 83 (4), 445–455.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S., Olson, M. V., 2000. Complete genome sequence of Pseudomonas aeruginosa pa01, an opportunistic pathogen. Nature 406 (6799), 959–964.

Valero, A., Perez-Rodriguez, F., Carrasco, E., Garcia-Gimeno, R. M., Zurera, G., 2006. Modeling the growth rate of Listeria monocytogenes using absorbance measurements and calibration curves. J. Food Sci. 71 (7), 257–M264.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., Riet, K. v., 1990. Modeling of the bacterial growth curve. Appl. Environ. Microbiol. 56 (6), 1875–1881.

PAPER

B

Maximum Likelihood based comparison of the specific growth rates for P. aeruginosa and four mutator strains [‡]

Abstract

The specific growth rate for *P. aeruginosa* and four mutator strains *mutT*, *mutY*, *mutM* and *mutY-mutM* is estimated by a suggested Maximum Likelihood, ML, method which takes the autocorrelation of the observation into account. For each bacteria strain, six wells of optical density, OD, measurements are used for parameter estimation. The data is log-transformed such that a linear model can be applied. The transformation changes the variance structure, and hence an OD- dependent variance is implemented in the model. The autocorrelation in the data is demonstrated, and a correlation model with an exponentially decaying function of the time between observations is suggested. A model with a full covariance structure containing OD-dependent variance and an autocorrelation structure is compared to a model with variance only and with no variance or correlation implemented. It is shown that the model that best describes data is a model taking into account the full covariance structure. An inference study is made in order to determine whether the growth rate of the five bacteria strains is the same. After applying a likelihood-ratio test to models with a full covariance structure, it is concluded that the specific growth rate is the same for all bacteria strains. This study highlights the importance of carrying out an explorative examination of residuals in order to make a correct parametrization of a model including the covariance structure. The ML method is shown to be a strong tool as it enables estimation of covariance parameters along with the other model parameters and it makes way for strong statistical tools for inference studies.

[‡]As published in: K. R. Philipsen, L. E. Christiansen, L. F. Mandsberg, O. Ciofu, H. Madsen, 2008. Maximum likelihood based comparison of the specific growth rates for *P. aeruginosa* and four mutator strains. Journal of Microbiological Methods 75 (3), 551-557

B.1 Introduction

Proper estimation of growth parameters is essential in many areas, for instance in determining the effect of antimicrobial treatment (Dalgaard and Koutsoumanis, 2001) or when modelling growth of bacteria in food processing and storage (Juneja et al., 2007; Shama et al., 2005). Furthermore, it is very important to be able to tell whether the growth of different bacteria strains is identical. This can form the basis of *in vivo* or *in vitro* experiments, such as competition experiments (Montanari et al., 2007), where two or more bacteria are competing to survive and overtake the population. If the growth rates of bacteria strains are not identical in a normal unstressed environment, this will affect the result of a competition experiment carried out in a stressed environment, e.g. by adding antibiotics. Thus, it is very important to correctly determine whether the growth rates are identical.

Bacterial growth is typically classified by the maximum growth rate $\mu_{\max}$ and the lag time (Baty and Delignette-Muller, 2004), when the growth rate is considered to be time dependent. Alternatively the growth is described by a Monod expression (Monod, 1949), which depends on the substrate content and contains the parameters $\mu_{\max}$ and the OD value were half the maximum growth is reached, κ_{50}. The Monod model should be considered when not enough substrate is available to reach intolerable numbers of bacteria before the growth rate decreases due to substrate depletion (Zwietering et al., 1990).

The objective of the current study is to determine whether the growth of *P. aeruginosa* and four mutator strains *mutT*, *mutY*, *mutM* and *mutY-mutM* can be regarded as identical. For this study optical density, OD, measurements are available for each strain growing in LB media. The study is motivated by a competition experiment between *P. aeruginosa* and each of the four mutator strains, for which interpretations of the results rely on the growth rates being identical. Examination of identical growth rates is relevant, as mutator strains are often considered to have lower fitness and thereby growth rate due to a higher mutation rate and thus more deleterious mutations. The mutation rates of the bacteria considered are listed in Table B.1. OD measurements are used in stead of CFU count, as this method demands less resources, and it is also the choice of measurement method for the competition experiment made. It has been argued (Augustin et al., 1999) that due to the detection limit of the OD measurements, the specific growth rate estimated for these OD values will be lower than the maximum specific growth. However, the specific growth rate is assumed to be a usable measure, for the purpose of determining whether the growth is the same for the five bacterial strains.

Recent studies (Baty and Delignette-Muller, 2004; Dalgaard and Koutsoumanis, 2001; Fujikawa et al., 2004; Juneja et al., 2007; Lindqvist, 2006; Shama et al., 2005) have compared the estimation of the growth rates and/or lag times obtained by different mathematical models. All of these studies use unweighted least squares to estimate the parameters. This paper suggests estimating the model parameters using the Maximum Likelihood (ML) method. This method enables us to

Table B.1: [a] Mandsberg et al. (2008), [b] calculated using the method described by Ma et al. (1992)

Bacteria strain	Mutation rate per generation	Ref.
P. aeruginosa	$4.61 \cdot 10^{-9}$	*a*
mutT	$1.28 \cdot 10^{-7}$	*a*
mutY	$3.85 \cdot 10^{-8}$	*a*
mutM	$6.38 \cdot 10^{-9}$	*a*
mutY-mutM	$1.94 \cdot 10^{-7}$	*b*

introduce a full model including a variance and autocorrelation structure for the observations and to determine the related parameters along with the growth parameters. The suggested ML method enables the use of strong statistical tools to compare models. As an example we apply the likelihood-ratio test to examine whether the growth of the five bacteria strains can be assumed to be identical. The study demonstrates the importance of including full information about variance and correlation structure in a growth model.

For the estimation of μ an exponential model is considered, which means that a linear model is fitted to the log-transformed OD curve where the slope is steepest (Zwietering et al., 1990). More advanced models have been proposed (for a recent review see Li et al. (2007)) to fit the growth curve, defined as the logarithm of the number of bacteria as a function of time. These models contain both lag phase and $\mu_{\max}$, so it is not necessary to subjectively decide the interval for the exponential growth. However these models are limited to sigmoidal growth curves, which do not describe well the growth of *P. aeruginosa* in LB media. Also the Monod model is not appropriate for describing growth on rich media (Kovárová-Kovar and Egli, 1998). Moreover, for the purpose of introducing a weighted estimation of the growth parameters, it is desirable and sufficient to keep the model as simple as possible. Therefore we consider the exponential model for growth.

B.2 Materials and Methods

B.2.1 Growth measurements

OD measurements are obtained for growth in LB media for *P. aeruginosa*, PAO1, and four different mutator strains; *mutT*, *mutY*, *mutM* and *mutY-mutM*. A description of the individual mutator strains is given in Mandsberg et al. (2008). The double mutant *mutY-mutM* is constructed by a method, not yet published, which is modified in accordance with Quenee et al. (2005). The bacteria are grown over night in LB media, after which they are adjusted to an OD600 on 0.03 and subsequently diluted to 10^{-4}. Each bacterium strain is transferred to six microtitter wells with $280\mu l$ to each well. Measurements are made with a sampling interval of 5 min in a bioscreen (Labsystem C (Bie og Berntsen)) at $37°C$ under continuous

shaking. The measurements are shown in Figure B.1.

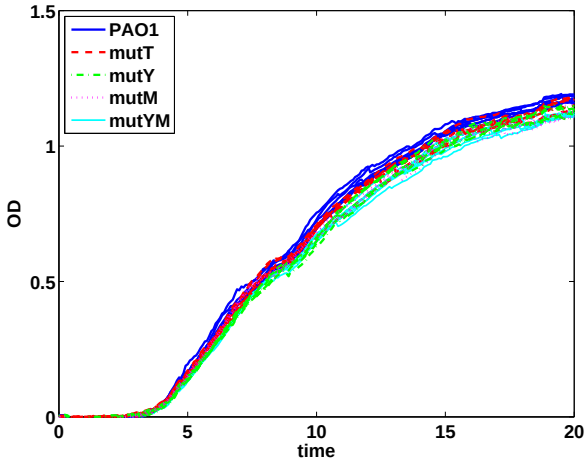


Figure B.1: The OD measurements for the five different bacteria, corrected for the OD of the media.

The specific growth rate occurs during exponential growth, and can be found by estimating the slope of the log-transformed data, where the slope is steepest. The interval for estimation of the specific growth rate shown in Figure B.2 is chosen by graphical inspection. The interval comprises 11 observations.

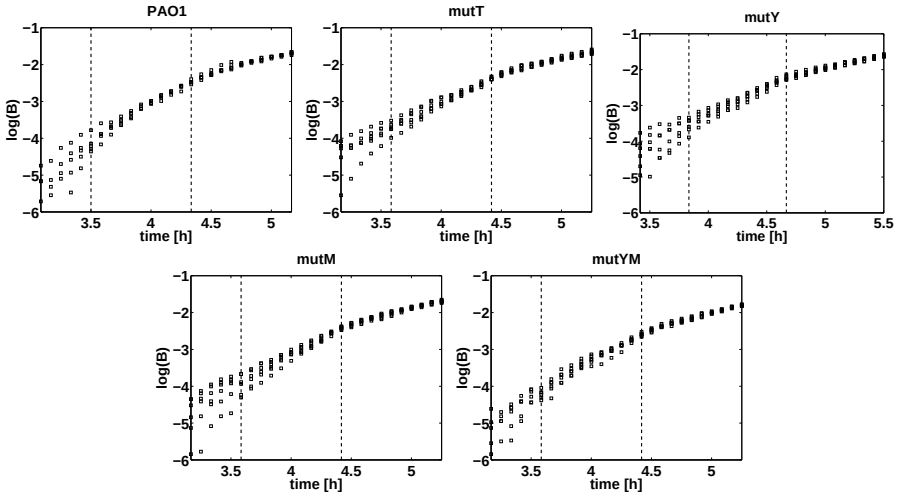


Figure B.2: The log-transformed OD measurements, corrected for media content, for the five different bacteria strains. The estimation interval for the maximal growth rate is marked by the vertical dotted lines.

Several authors (Chorin et al., 1997; Augustin et al., 1999; Baty et al., 2002) have discussed the relation between OD and CFU measurements and the influence of

the measurement method on the estimated growth parameters. For this study the relation between OD and the actual bacterial concentration has been examined experimentally, and the relation between OD and concentration is found to be linear in the examined interval. Since the relation is linear a transformation of the data from OD to CFU will not influence the estimate of the specific growth rate, and therefore the following study is continued with the OD values.

B.2.2 Model

The OD measurements are initially transformed by

$$\begin{aligned} B_{bij} &= \mathrm{OD}_{bij} - \mathrm{M}_j \\ Y_{bij} &= \log(B_{bij}), \end{aligned} \tag{B.1}$$

where OD_{bij} is the measured OD value for bacteria strain b ($b = 1,2,\ldots,S$), repetition i ($i = 1,2,\ldots,R$) at time t_j ($j = 1,2,\ldots,T$). M_j is the mean of OD values for ten wells with media without bacteria, and thus B_{bij} is the OD contribution due to growth, corrected for the media. In this study $S = 5$, $R = 6$ and $T = 11$.

The linear relation seen in Figure B.2 for the log-transformed data is modelled by a general linear model of the form

$$\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{\epsilon} \ , \text{where } \boldsymbol{Y} \in N(\boldsymbol{X\theta}, \sigma^2\boldsymbol{\Sigma}) \, , \tag{B.2}$$

where $\boldsymbol{X}$ is the design matrix and $\boldsymbol{\theta}$ is a set of parameters $[\boldsymbol{\alpha}, \boldsymbol{\mu}]$ with $\boldsymbol{\alpha}$ being the intercept and $\boldsymbol{\mu}$ the slope, i.e. the specific growth rate. $\boldsymbol{Y}$ is a vector of length SRT containing all the observations in the estimation interval. To reduce the correlation between the estimated values for $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ the time series are translated such that they starts at $t_1 = 0$. The values contained in of $\boldsymbol{\alpha}$ is thus the $\boldsymbol{Y}$ values at the beginning of the estimation interval.

Two models are introduced with the purpose of determining whether all bacteria strains can be assumed to have the same specific growth: Model 1 where the growth rate is different for each bacteria strain, and Model 0 where the growth rate is the same for all repetitions of the experiment. For both models the intercept is different for all bacteria in order to account for the small difference in initial OD and media concentration. The following notation will be used for the vector of all time points $\boldsymbol{t} = [t_1, t_2, \ldots, t_T]^{\mathrm{T}}$, and $\boldsymbol{T}$ is a column vector with R repetitions of the vector $\boldsymbol{t}$. A column vector of length T containing only ones is written in short-hand notation as $\boldsymbol{e} = [1, 1, \ldots, 1]^{\mathrm{T}}$ and a matrix comprising R repetitions of $\boldsymbol{e}$ is defined as

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{e} & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{e} & & \vdots \\ \vdots & & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{e} \end{bmatrix} .$$

Model 1:

$$Y_{bij}^{(1)} = \alpha_{bi} + \mu_b t_j + \epsilon_{bij}\,, \tag{B.3}$$

or in matrix formulation

$$\begin{aligned} \boldsymbol{Y}^{(1)} &= \boldsymbol{X}_1\boldsymbol{\theta}_1 + \boldsymbol{\epsilon}\,, \text{ where} \\ \boldsymbol{X}_1 &= \begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{T} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{E} & & \vdots & \boldsymbol{0} & \boldsymbol{T} & & \vdots \\ \vdots & & \ddots & \boldsymbol{0} & \vdots & & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{E} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{T} \end{bmatrix}, \\ \boldsymbol{\theta}_1 &= \left[\alpha_{11},\dots,\alpha_{1R},\alpha_{21},\dots,\alpha_{SR},\mu_1,\dots,\mu_S\right]^{\mathrm{T}} \end{aligned} \tag{B.4}$$

Model 0:

$$Y_{bij}^{(0)} = \alpha_{bi} + \mu t_j + \epsilon_{bij}\,, \tag{B.5}$$

or in matrix formulation

$$\begin{aligned} \boldsymbol{Y}^{(0)} &= \boldsymbol{X}_0\boldsymbol{\theta}_0 + \boldsymbol{\epsilon}\,, \text{ where} \\ \boldsymbol{X}_0 &= \begin{bmatrix} \boldsymbol{e} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{t} \\ \boldsymbol{0} & \boldsymbol{e} & & \vdots & \boldsymbol{t} \\ \vdots & & \ddots & \boldsymbol{0} & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{e} & \boldsymbol{t} \end{bmatrix}, \\ \boldsymbol{\theta}_0 &= \left[\alpha_{11},\alpha_{12},\dots,\alpha_{1R},\alpha_{21},\dots,\alpha_{SR},\mu\right]^{\mathrm{T}} \end{aligned} \tag{B.6}$$

When introducing Model 1 it is assumed that the specific growth rate for each repetition within a bacterium strain is the same. This is biologically plausible as the bacteria and media mixture in each of the six wells come from the same batch culture. A third model has been considered where the specific growth rate for each repetition is different, but the analysis indicates that there are too few data points for each curve to give a good estimation of the model parameters. Therefore the rest of this study continues with Model 0 and Model 1.

B.2.3 Maximum Likelihood estimation

The model parameters are estimated by maximizing the log-likelihood function. The likelihood function is equal to the joint probability density of the data, $p(\boldsymbol{Y}|\boldsymbol{\theta})$

$$L(\boldsymbol{\theta}|\boldsymbol{Y}) = p(\boldsymbol{Y}|\boldsymbol{\theta})\,. \tag{B.7}$$

As the data can be assumed to be normally distributed, the probability density for $\boldsymbol{Y}$ is

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^N \det\boldsymbol{\Sigma}}} \cdot \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\theta})\right), \tag{B.8}$$

where N is the total number of observations. The log-likelihood function for normally distributed data is thus

$$\begin{aligned}\log(L(\boldsymbol{\theta}|\boldsymbol{Y})) = &-\frac{1}{2}N\log(\sigma^2) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) \\ &- \frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\theta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X\theta})\end{aligned} \tag{B.9}$$

plus a constant term $(-\frac{1}{2}\log(2\pi))$, which for simplicity is ignored, since it does not depend on the parameters.

In order to parameterize $\boldsymbol{\Sigma}$, an examination of the variance and autocorrelation structure of the data is needed. Assuming that the variance of $\boldsymbol{B}$ is Var[$\boldsymbol{B}$], the variance of the log-transformed data can be determined from

$$\text{Var}[f(x)] = \text{Var}[x]f'(x)^2 \tag{B.10}$$

i.e.,

$$\text{Var}[\boldsymbol{Y}] = \text{Var}[\boldsymbol{B}]\left(\frac{1}{\boldsymbol{B}}\right)^2 = \sigma^2\left(\frac{1}{\boldsymbol{B}}\right)^2. \tag{B.11}$$

Indeed the variance depends on the inverse of the square of $\boldsymbol{B}$, as seen in Figure B.3. The figure shows the variance of $\boldsymbol{Y}$ as calculated from six repetitions within each bacteria strain at each time plotted, together with a fit of the theoretical expressions in equation (B.11) to the inverse square mean of $\boldsymbol{B}$.

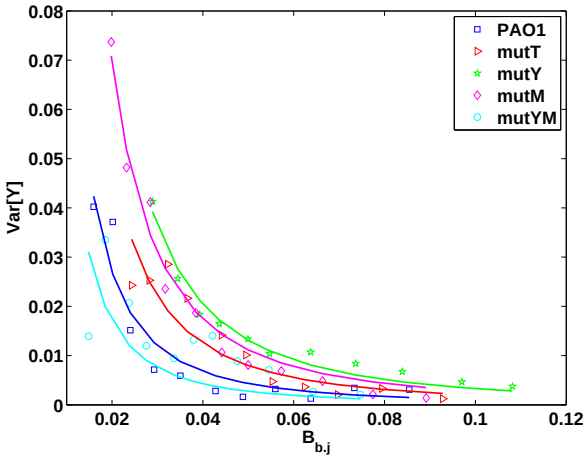


Figure B.3: The variance of $\boldsymbol{Y}$ against $B_{b.j}$ as calculated from data (symbols) plotted together with $\sigma^2(1/B_{b.j})^2$ (lines), where $B_{b.j}$ is the mean value for each time point and bacteria strain and σ^2 is estimated by least squares regression.

A significant autocorrelation was found in earlier studies (López et al., 2004), and should be included in the model to give a full parameterization of the data. In order to determine the correlation structure of the residuals ϵ_{bij}, a ML estimate of Model 0 and Model 1 with $\boldsymbol{\Sigma} = I$ is initially examined by plotting the autocorrelation function in Figure C.5. This plot indicates that the noise sequence is indeed

correlated in time. For simplicity we suggest the following exponentially decaying function of the time between two observations

$$\mathrm{Corr}[\epsilon_j, \epsilon_k] = \rho^{|j-k|} \tag{B.12}$$

to describe the autocorrelation (Madsen and Thyregod, 1988). This parametrization of the autocorrelation can be chosen since the observations are equidistant. The parameter ρ now corresponds to the lag 1 correlation, i.e. the correlation between two consecutive observations.

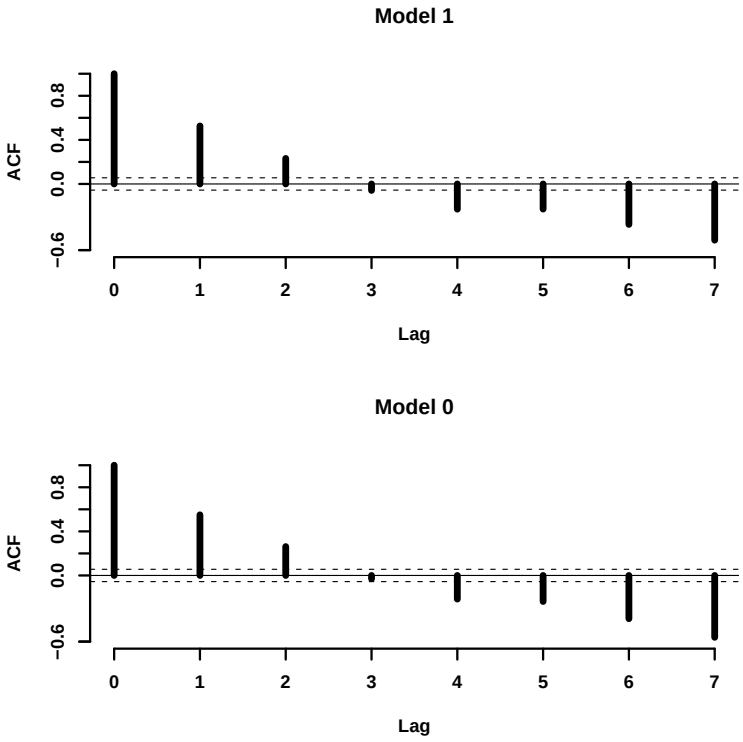


Figure B.4: The autocorrelation function, ACF, for the residuals of Model 0 and Model 1 with $\boldsymbol{\Sigma}$ being the identity matrix fitted to the measured OD values.

Three different structures of the covariance matrix $\boldsymbol{\Sigma}$ are examined in order to compare the models and determine the $\boldsymbol{\Sigma}$ that best describes data. The most simple is the identity matrix

$$I: \ \boldsymbol{\Sigma} = \boldsymbol{I}\,, \tag{B.13}$$

for which the ML estimation corresponds to performing a least squares estimate of the model parameters. On the basis of the theoretical explanation and explorative examination of the data, it is clear that the variance depends on $\boldsymbol{B}$ as described in

(B.11). Therefore the following variance structure for each bacterium b and each repetition i is suggested

$$II: \left\{\Sigma_{jj}^{bi}\right\} = \frac{1}{(B_{bij})^2}\,, \tag{B.14}$$

where B_{bij} is the OD measurement of repetition i at time k for bacteria b less the contribution from the media M_j.

Furthermore, in order to include the autocorrelation as well as the variance structure, we introduce the full $\boldsymbol{\Sigma}$ matrix. This is a block diagonal matrix with one block matrix for each repetition of the experiment, thus for each bacterium b and repetition i it is given by

$$III: \left\{\Sigma_{jk}^{bi}\right\} = \frac{\rho^{|j-k|}}{B_{bij}B_{bik}} \tag{B.15}$$

where Σ_{jk}^{bi} is the jk element ($j = 1,2,...,T$ and $k = 1,2,...,T$) of the block matrix belonging to repetition i and bacteria b.

The total set of parameters to be estimated is thus σ, ρ and $\boldsymbol{\theta}$ where θ_{bij} contains the model parameters α_{bij} and μ_{bij}. In order to reduce the computation time for the estimation, only the parameters σ and ρ are estimated by nonlinear optimization. With these parameters given, the remaining model parameters $\boldsymbol{\theta}$ are found by the ML optimization as

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\,, \tag{B.16}$$

where $\boldsymbol{\Sigma}$ equals one of the expressions (B.13), (B.14) or (B.15). The resulting parameter estimation is equivalent to estimating all parameters simultaneously.

The variance of the estimated parameters and the correlation between the estimated parameters can be calculated from the inverse Hessian, where the Hessian matrix H is equal to the second order partial derivative of the log-likelihood function, $\ell = \log(L(\boldsymbol{\theta}|\boldsymbol{Y}))$ (B.9). Derivation of the Hessian matrix is found in Appendix B.A.

The models, the log-likelihood function and the algorithm for calculating the Hessian matrix are implemented in `Matlab 7.3.0 (R2006b)`. The `Matlab` command `fminsearch` is used to determine the maximum of the log-likelihood function.

B.2.4 Likelihood-ratio test

The likelihood-ratio test is used for an inference study concerning the nested models (B.3) and (B.5). The hypothesis is that the specific growth rate for each bacterial population is identical, and it can thus be described by Model 0. This hypothesis is biological plausible if the possibility of mutations is low within the considered interval. The test statistic is

$$-2\log\left(\frac{L_0}{L_1}\right) = -2\left(\ell_0 - \ell_1\right)\,, \tag{B.17}$$

which is asymptotically χ^2 distributed with degrees of freedom corresponding to the difference in number of parameters between the two models tested. Here ℓ_0 is the log-likelihood value for Model 0, and ℓ_1 is the log-likelihood value for Model 1,- both evaluated at their optimal value.

B.3 Results and discussion

The model parameters have successfully been estimated for Model 1 (B.3) and Model 0 (B.5) for each of the three proposed covariance matrices I (B.13), II (B.14) and III (B.15). Residuals from the estimation have been examined in Figure B.5 where the residuals are plotted as a function of index. In the first case, where no variance structure is introduced in the model, an unclear residual structure is seen. For the covariance matrices II and III, the variance of the residuals follows the estimated squared structure, which indicates that the variance has been implemented correctly. The correlation structure is examined by plotting the au-

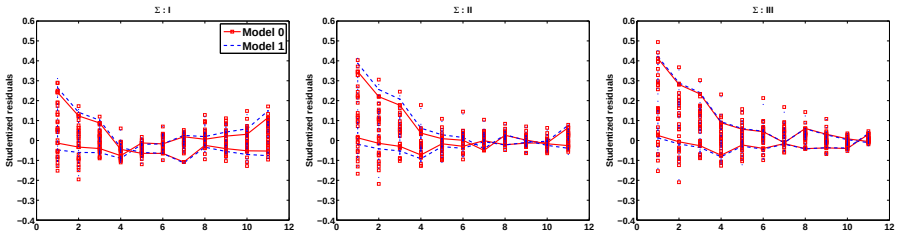


Figure B.5: Residuals for the ML estimation of Model 0 and Model 1. Two curves are shown for PA01 and mutY, respectively, to illustrate the correlation structure in the data.

tocorrelation of $\boldsymbol{\epsilon}$ for the full covariance structure in Figure B.6. An exponentially decaying correlation is observed, as initially assumed. Thus, there seems to be no unexplained variance or correlation structure when applying the suggested full covariance matrix.

Table B.2: Estimates of the specific growth rate μ and doubling time T_d.

	Growth rate [1/h] (SD)			T_d [min]
$\boldsymbol{\Sigma}$ form	I	II	III	III
PAO1	1.77 (0.032)	1.653 (0.023)	1.566 (0.045)	26.56
mutT	1.57 (0.032)	1.524 (0.023)	1.517 (0.039)	27.42
mutY	1.64 (0.032)	1.534 (0.028)	1.528 (0.049)	27.22
mutM	1.74 (0.032)	1.644 (0.025)	1.560 (0.047)	26.66
mutYM	1.74 (0.032)	1.657 (0.030)	1.621 (0.053)	25.66
All	1.69 (0.015)	1.600 (0.012)	1.546 (0.026)	26.90

The growth rates for each bacteria type estimated by Model 1 and Model 0, are listed in Table B.2, for each of the three suggested covariance matrices I, II

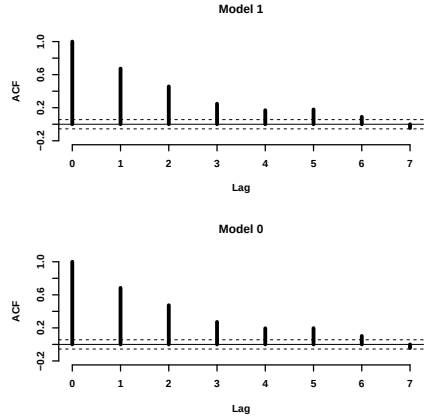


Figure B.6: Autocorrelation function for Model 0 and Model 1 with the full $\mathbf{\Sigma}$ matrix.

and III. The growth rate is generally estimated highest when $\mathbf{\Sigma}_I$ is used, and lowest when the full covariance matrix is applied. The difference in the estimated specific growth rates shows that it is very important to use the correct covariance matrix in order to obtain a correct estimation for the growth rates. In many studies the doubling time T_d is used instead of the growth rate. The doubling time and growth rate are related by $T_d = \log(2)/\mu$. The doubling time for the full model is given in Table B.2 to assist comparison with other microbiological studies. It should be noted that the specific growth rate might be smaller than the maximum specific growth rate, as explained in the introduction. It would therefore be of interest for a future study to repeat the estimation with a CFU count experiment, in order to examine the difference between the two experimental methods.

Table B.3: Estimates of the standard deviation and correlation parameters σ and ρ.

	$\mathbf{\Sigma}$ form	Model 1	Model 0
σ(SD)	I	0.0685 (0.0027)	0.0715 (0.0028)
	II	0.0026 (0.0001)	0.0027 (0.0001)
	III	0.0029 (0.0003)	0.0031 (0.0003)
ρ(SD)	III	0.7360 (0.0553)	0.7656 (0.0550)

The log-likelihood values for each model are given in Table B.4. The models with $\mathbf{\Sigma}_{II}$ and $\mathbf{\Sigma}_{III}$ can be compared using the likelihood-ratio test, as these models are nested. Doing this for Model 1 gives a test statistic of 148.62, which using a χ^2 distribution with one degree of freedom gives a P value close to 0. The same result is obtained by Model 0. This means that $\mathbf{\Sigma}_{III}$ should be used instead of $\mathbf{\Sigma}_{II}$ and indicates further that the full covariance matrix is preferable to the identity matrix $\mathbf{\Sigma}_I$. This conclusion is further emphasized by the estimated value for ρ shown in Table B.3, which clearly shows that there is autocorrelation in the data, and this

should therefore be included in the model to give a correct estimate for μ. The same table shows the estimates for σ. As expected, the variance is higher when the identity matrix $\mathbf{\Sigma}_I$ is used, than when the increased variation for smaller $\boldsymbol{B}$ values is accounted for.

The correlations between the estimated parameters are obtained from the inverse Hessian matrix. If the time series are not translated as described in Section B.2.2 a very high correlation on up to 0.999 is seen between the intercept and the related maximal growth rate. By translating the time series, this correlation is reduced to 0.665 for Model 1 and 0.468 for Model 0, which is why the translated time series are used for the estimation. For the translated data none of the correlations of the estimated parameters are critical. The highest correlation is found to be between σ and ρ (0.901 for Model 1 and 0.922 for Model 0).

Table B.4: The results of the inference analysis.

$\mathbf{\Sigma}$ form	$\log(L_1)$	$\log(L_0)$	P
I	719.69	705.74	$1.30 \cdot 10^{-5}$
II	799.94	786.18	$1.56 \cdot 10^{-5}$
III	874.25	872.89	0.607

The results of the inference analysis for the two models are summarized in Table B.4. The result is very dependent on the choice of $\mathbf{\Sigma}$. For $\mathbf{\Sigma}_I$ and $\mathbf{\Sigma}_{II}$ the specific growth rate for the bacteria strains cannot be regarded as the same. However, we have argued that the full covariance matrix $\mathbf{\Sigma}_{III}$ must be used for the model to describe the total variance and correlation structure of the data. For the full covariance matrix, it can be concluded from the inference study that the specific growth rate is the same for all bacteria strains.

The different results for the three different $\mathbf{\Sigma}$ highlight the importance of including the correct covariance structure in the model. In this connection the ML estimation is preferable to least squares estimation, as the parameters for $\mathbf{\Sigma}$ can be estimated along with the other model parameters.

The model examined in this study is a linear model, but it can easily be replaced by a non-linear model (Madsen and Thyregod, 1988; Madsen, 2008). The disadvantage of introducing a non-linear model is that this significantly increases the complexity of the Hessian matrix, so that it might not be possible to calculate it analytically. However many computational tools are available for calculating the Hessian matrix numerically. Therefore, a continuation of this study could be to introduce a non-linear model which can describe the entire growth process. This would require the use of a substrate-dependent growth function as the growth enters the stationary phase due to substrate depletion.

B.4 Conclusions

The importance of including full variance and correlation structure in a model for bacterial growth has been shown. The estimation of model parameters is dependent on the parametrization of the covariance matrix, and disregarding the variance and correlation structure can therefore have consequences for the results of a study.

In this study the objective was to estimate the specific growth rate for PAO1 and four mutator strains. An explorative analysis of the OD measurements showed a strong correlation in time. The correlation was successfully described by an exponentially decaying function of the time between observations. Additionally, a variance structure for the log-transformed observations was implemented. An ML approach to estimating the model parameters is used. As an example of the strong statistical tools available with the ML method, we use the likelihood-ratio test to determine whether the growth rates of the five bacteria strains can be assumed to be identical. From the test it can be concluded that the specific growth rate is indeed the same.

Acknowledgements

This study was supported by the Danish Research Council for Technology and Production Sciences through the project "Evolution and adaptation of antimicrobial resistance in bacterial populations".

B.5 Bibliography

Augustin, J.-C., Rosso, L., Carlier, V., 1999. Estimation of temperature dependent growth rate and lag time of listeria monocytogenes by optical density measurements. J. Microbiol. Methods 38 (1-2), 137–146.

Baty, F., Delignette-Muller, M.-L., 2004. Estimating the bacterial lag time: which model, which precision? Int. J. Food Microbiol. 91, 261–277.

Baty, F., Flandrois, J. P., Delignette-Muller, M. L., 2002. Modeling the lag time of *Listeria monocytogenes* from viable count enumeration and optical density data. Appl. Environ. Microbiol. 68, 5816–5825.

Chorin, E., Thuault, D., Cléret, J.-J., Bourgeois, C.-M., 1997. Modelling bacillus cereus growth. Int. J. Food Microbiol. 38 (2-3), 229–234.

Dalgaard, P., Koutsoumanis, K., 2001. Comparison of maximum specific growth rates and lag times estimated from absorbance and viable count data by different mathematical models. J. Microbiol. Methods 43 (3), 183–196.

Fujikawa, H., Kai, A., Morozumi, S., 2004. A new logistic model for *Escherichia coli* growth at constant and dynamic temperatures. Food Microbiol. 21 (5), 501–509.

Juneja, V., Valenzuela Melendres, M., Huang, L., Gumudavelli, V., Subbiah, J., Thippareddi, H., 2007. Modeling the effect of temperature on growth of *Salmonella* in chicken. Food Microbiol. 24 (4), 328–335.

Kovárová-Kovar, K., Egli, T., 1998. Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics. Microbiol. Mol. Biol. Rev. 62 (3), 646–666.

Li, H., Xie, G., Edmondson, A., 2007. Evolution and limitations of primary mathematical models in predictive microbiology. Brit. Food J. 109 (8), 608–626.

Lindqvist, R., 2006. Estimation of *Staphylococcus aureus* growth parameters from turbidity data: Characterization of strain variation and comparison of methods. Appl. Environ. Microbiol. 72 (7), 4862–4870.

López, S., Prieto, M., Dijkstra, J., Dhanoa, M., France, J., 2004. Statistical evaluation of mathematical models for microbial growth. Int. J. Food Microbiol. 96 (3), 289–300.

Ma, W. T., Sandri, G. V., Sarkar, S., 1992. Analysis of the luria-delbrück distribution using discrete convolution powers. J. Appl. Prob. 29 (2), 255–267.

Madsen, H., 2008. Time series analysis. Chapman & Hall/CRC.

Madsen, H., Thyregod, P., 1988. Modelling the time correlation in hourly observations of direct radiation in clear skies. Energ. Buildings 11, 201–211.

Mandsberg, L. F., Ciufo, O., Christensen, L. E., Højby, N., 2008. Antibiotic resistance in hypermutable *P. aeruginosa* due to inactivation of the dna oxidative repair system, submitted to: Antimicrobial Agents and Chemotherapy.

Monod, J., 1949. The growth of bacterial cultures. Annu. Rev. Microbiol. 3, 371–394.

Montanari, S., Oliver, A., Salerno, P., Mena, A., Bertoni, G., Tummler, B., Cariani, L., Conese, M., Doring, G., Bragonzi, A., 2007. Biological cost of hypermutation in *Pseudomonas aeruginosa* strains from patients with cystic fibrosis. Microbiology - Reading 153 (5), 1445.

Quenee, L., Lamotte, D., Polack, B., 2005. Combined sacb-based negative selection and cre-lox antibiotic marker recycling for efficient gene deletion in pseudomonas aeruginosa. BioTechniques 38 (1), 63–67.

Shama, G., Perni, S., Andrew, P., 2005. Estimating the maximum growth rate from microbial growth curves: definition is everything. Food Microbiol. 22 (6), 491–5.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., Riet, K. v., 1990. Modeling of the bacterial growth curve. Appl. Environ. Microbiol. 56 (6), 1875–1881.

B.A Derivation of the Hessian matrix

For the full model containing the covariance matrix (B.15) the Hessian is

$$H = \begin{bmatrix} \ell_{\sigma,\sigma} & \ell_{\sigma,\rho} & \ell_{\sigma,\boldsymbol{\theta}} \\ \ell_{\rho,\sigma} & \ell_{\rho,\rho} & \ell_{\rho,\boldsymbol{\theta}} \\ \ell_{\boldsymbol{\theta},\sigma} & \ell_{\boldsymbol{\theta},\rho} & \ell_{\boldsymbol{\theta},\boldsymbol{\theta}} \end{bmatrix}, \tag{B.18}$$

where ℓ_{p_1,p_2} is the second order derivative of ℓ with respect to the parameters p_1 and p_2. Before continuing some short-hand notation is introduced

$$g(\rho) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{\Sigma}(\rho)^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}) \text{ , and} \tag{B.19}$$

$$\boldsymbol{\epsilon} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\theta}} \tag{B.20}$$

The first order partial derivatives for ℓ are

$$\ell_{\boldsymbol{\theta}} = \frac{1}{\sigma^2}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right) \tag{B.21}$$

$$\ell_{\sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}g(\rho) \tag{B.22}$$

$$\ell_{\rho} = -\frac{1}{2}\mathrm{Tr}\Big[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho} \tag{B.23}$$

$$+ \frac{1}{2\sigma^2}\left(\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right)\Big] \tag{B.24}$$

where

$$\boldsymbol{\Sigma}_{\rho}(jk) = \frac{\partial \Sigma^{bi}_{jk}}{\partial \rho} = |j-k|\frac{\rho^{|j-k|-1}}{B_{bij}B_{bik}} \tag{B.25}$$

The second order partial derivatives of ℓ are

$$\ell_{\boldsymbol{\theta},\boldsymbol{\theta}} = -\frac{1}{\sigma^2}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) \tag{B.26}$$

$$\ell_{\sigma,\sigma} = \frac{N}{\sigma^2} - \frac{3}{\sigma^4}g(\rho) \tag{B.27}$$

$$\ell_{\rho,\rho} = -\frac{1}{2}\mathrm{Tr}\Big[-\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho,\rho}\Big] + \frac{1}{2\sigma^2}\Big[\boldsymbol{\epsilon}^{\mathrm{T}}\Big[-2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho,\rho}\boldsymbol{\Sigma}^{-1}\Big]\boldsymbol{\epsilon}\Big] \tag{B.28}$$

$$\ell_{\boldsymbol{\theta},\boldsymbol{\theta}} = -\frac{1}{\sigma^2}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) \tag{B.29}$$

$$\ell_{\sigma,\rho} = -\frac{1}{\sigma^3}\left[\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right] \tag{B.30}$$

$$\ell_{\boldsymbol{\theta},\sigma} = -\frac{2}{\sigma^3}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right) \tag{B.31}$$

$$\ell_{\boldsymbol{\theta},\rho} = -\frac{1}{\sigma^2}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\rho}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right) \tag{B.32}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{\rho,\rho}(jk) &= \frac{\partial^2 \Sigma^{bi}_{jk}}{\partial \rho^2} \\ &= |j-k|(|j-k|-1)\frac{\rho^{|j-k|-2}}{B_{bij}B_{bik}}\end{aligned} \tag{B.33}$$

PAPER

C

Modelling bacterial growth in rich media with a non-parametric extension to an SDE based model ‡

Abstract

In this study a systematic framework for model improvement based on stochastic differential equations has been presented and applied to examine the growth of bacteria in rich media. The modelling framework uses available data to dig into embedded information of the dynamic of the system by extending the stochastic differential equation model with a non-parametric state. The case considered is the growth of *Salmonella* and *Enterococcus* in rich media. It is found to be necessary to include an extra state in the model to capture the lag phase of the growth. Furthermore it is shown that the Monod equation, which has traditionally been used to describe bacterial growth kinetics, can be replaced by a growth rate proportional to the substrate level.

C.1 Introduction

A good description of bacterial growth kinetics, i.e. the relationship between bacterial growth and substrate concentration is important for many applications in microbiology, for instance for fermentation processes (Kompala et al., 1984; Patnaik, 1999). The Monod equation was the first suggestion for a mathematical description of the growth curve. It has been extensively discussed since its introduction in 1949 (Monod, 1949). When originally proposed, it seemed to be a "convenient and logical" (Monod, 1949) choice for a mathematical expression to fit the growth curve. Several attempts have been made to formulate a mechanistic

‡Manuscript under preparation: K. R. Philipsen, L. E. Christiansen, H. Madsen, Modelling bacterial growth in rich media with a non-parametric extension to an SDE based model. Part of this manuscript is included in: J. K. Møller, K. R. Philipsen, L. E. Christiansen, H. Madsen, 2010. Analysis of diffusion in an SDE-based bacterial growth model. Submitted.

background for the Monod equation (Lobry et al., 1992; Liu, 2006). Even though the Monod equation is a good estimation of the growth on a single substrate, it fails to model growth on rich media (Bajpai-Dikshit et al., 2003). Therefore several attempts have been made to find another equation for this growth (Doshi et al., 1997). According to Kovárová-Kovar and Egli (1998) these studies can be divided into three methods i) Extending the Monod model with additional constants, ii) Developing an empirical or mechanistic model from kinetic concepts, iii) Describing how the Monod growth parameters are influenced by physicochemical factors. A problem with determining other expressions for the growth rate has been the lack of high quality reproducible data which relates the growth rate to the substrate concentration (Kovárová-Kovar and Egli, 1998). The method proposed here makes it possible to extract these data from bioscreen measurements, and thus limit the time and resources used on experiments significantly.
The transition from modelling using Ordinary Differential Equations (ODEs) to using Stochastic Differential Equations (SDEs) paves the way for many strong statistical tools for model development and inference (Philipsen et al., 2010a). In this article a model development method based on SDEs will be introduced and used to examine the relation between growth rate and substrate concentration. The use of SDEs enables the separation of measurement noise and system noise, which will be used in the method. First the SDE framework will be introduced, followed by an outline of the suggested methods for model development. In order to demonstrate the method on a system with a known growth expression a simulation study is performed. Finally the model development method is used to study the growth expression for *Salmonella* and *Enterococcus* when growing in BHI media.

C.2 Modelling using Stochastic Differential Equations

The method for model building introduced in this article is based on SDE model formulations. A widely used (Kristensen et al., 2004b) continuous-discrete time SDE-based state space model consists of a continuous time system equation (the SDE) and a discrete time measurement equation, i.e.

$$d\boldsymbol{X_t} = \boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t\,, \tag{C.1}$$

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k\,, \tag{C.2}$$

where Equation (C.1) is the system equation and Equation (C.2) is the observation equation. $\boldsymbol{X}_t$ is a vector of state variables, $\boldsymbol{Y}_k$ is a vector of measured output variables, $\boldsymbol{u}_t$ is a vector of known input variables, $\boldsymbol{e}_k$ is an l-dimensional white noise process with $\boldsymbol{e} \in \mathcal{N}(0, \boldsymbol{R}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}))$ and $\boldsymbol{\omega}_t$ is a standard Wiener process with zero mean and independent Gaussian time increments with diffusion coefficient $\boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})$. The first part of the system equation is called the drift term and the second part is called the diffusion term. The Wiener process adds noise to the sys-

tem, more specifically $d\omega_t$ has uncorrelated random increments. Everywhere in this article Itô interpretations of the SDEs are used (Øksendal, 2007).

C.2.1 Case: Bacterial growth on rich media

Bacterial growth is traditionally modelled using deterministic equations such as a set of coupled Ordinary Differential Equations (ODEs)

$$\begin{aligned} \frac{dB}{dt} &= \mu(S,\boldsymbol{\theta})B\,, \\ \frac{dS}{dt} &= -\eta\mu(S,\boldsymbol{\theta})B\,, \end{aligned} \tag{C.3}$$

where B and S are the concentrations of the bacteria and the substrate, respectively, η is the amount of substrate necessary for creation of a new bacteria by cell division and $\mu(S,\boldsymbol{\theta})$ is the substrate-dependent growth rate with parameters $\boldsymbol{\theta}$. $1/\eta$ is in many studies called the biomass yield coefficient or yield factor. Given that Equation (C.3) is used to describe the system, η can be calculated from data as

$$\frac{dS}{dt} = -\eta\frac{dB}{dt} \Rightarrow \tag{C.4}$$

$$\int_{S_0}^{S_{\min}} dS = \int_{B_0}^{B_{\max}} \eta dB \Rightarrow \tag{C.5}$$

$$\eta = \frac{S_0 - S_{\min}}{B_{\max} - B_0} \tag{C.6}$$

For this study S is implemented in the model as a standardized substrate content, i.e. $S \in [0,1]$. η can therefore be estimated from

$$\eta = \frac{1}{B_{\max} - B_0}\,, \tag{C.7}$$

given that all the substrate has been used in the experiment as the bacteria concentration reaches maximum. The most frequently used mathematical equation to describe the kinetics of bacterial growth is the Monod equation (Monod, 1949)

$$\mu(S) = \frac{\mu_{\max} S}{\kappa + S}\,. \tag{C.8}$$

It is straightforward to transfer the well known ODE description of bacterial growth into the SDE framework. The right-hand side of the ODE (C.3) corresponds to the drift part $\boldsymbol{f}(\cdot)$ of the SDE (C.1). The ODE can thereby be translated into an SDE model with the state equation

$$d\begin{bmatrix} B_t \\ S_t \end{bmatrix} = \begin{bmatrix} \mu(S_t,\boldsymbol{\theta})B_t \\ -\eta\mu(S_t,\boldsymbol{\theta})B_t \end{bmatrix} dt + \begin{bmatrix} \sigma_B u_t & 0 \\ -\eta\sigma_B u_t & 0 \end{bmatrix} d\boldsymbol{\omega}_t\,. \tag{C.9}$$

In this study a state noise proportional to the observed bacterial concentration is used, as the variance of the system can be assumed to increase with increasing

bacterial concentration. That is, when estimating the model parameters the input u_t consists of the observed bacterial concentration Y_k, and when the model is simulated u_t is replaced by the state variable B_t. Due to mass balance, an increase in the bacterial concentration driven by the diffusion term also decreases the substrate content. Therefore, the term $\eta\sigma_B B$ is subtracted in the diffusion term for the substrate.
The bioscreen only measures the bacterial concentration and not the substrate concentration, hence the measurement equation is

$$Y_k = B_k + e_k\,,\; e_k \in \mathcal{N}(0,R)\,, \tag{C.10}$$

where e_k is the measurement error at time t_k and the measurement variance R is assumed to be constant for all times.
The software `CTSM` (`www.imm.dtu.dk/~ctsm`) is used to estimate parameters, and generate smoothing and prediction data. CTSM is to our knowledge unique in its ability to estimate model parameters and generate smoothing data for non-linear SDEs observed with measurement noise at discrete points in time. A user guide as well as detailed mathematical description of the available tools can be found on the software website.

C.3 Method

Modelling a dynamical system often consists of several steps of parameter estimation and model validation. If a model does not give a good fit to data, different methods for model improvement may be used. In this paper a new method for systematic model improvement extended from the framework by Kristensen et al. (2004a) will be described and applied to study the dynamics of bacterial growth. The steps constituting the method are shown in Figure C.1. The theory involved in each step will be elaborated on in relation to the bacterial growth study in the following subsections.

C.3.1 Part I (parametric)

Model specification - Step Ia

The first step when modelling a dynamical system is to suggest one or more candidate models which can be tested against data. It is advisable to examine the literature for existing models to use as candidate models and to use available physical insight about the system.
As a first candidate model for the bacterial growth we will use the SDE model (C.9) and the widely used Monod equation (C.8) to model the growth kinetics.

Parameter estimation - Step Ib

The parameters in the SDE-based model (C.1)-(C.2) are estimated by a maximum likelihood, ML, method (Kristensen et al., 2004b) using the software `CTSM`. The

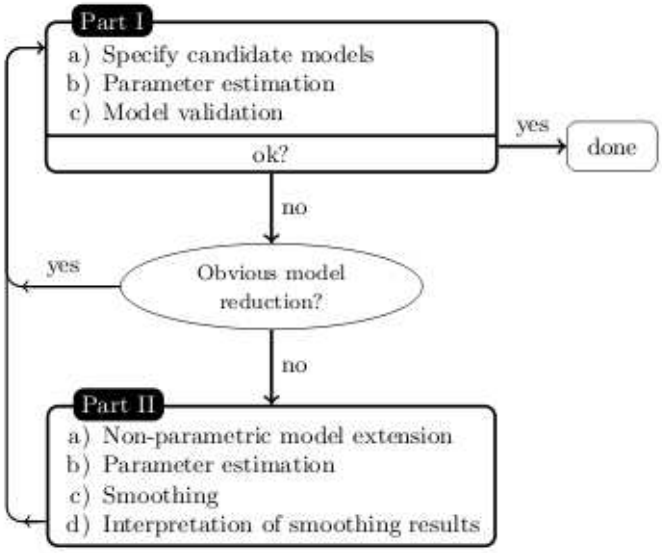


Figure C.1: Schematic overview of a systematic framework for model improvement.

unknown model parameters can be found by maximizing the likelihood function

$$L(\boldsymbol{\theta};\mathscr{Y}_N) = \left(\prod_{k=1}^{N} p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta})\right) p(\boldsymbol{Y}_0|\boldsymbol{\theta}) \tag{C.11}$$

with respect to a known sequence of observations

$$\mathscr{Y}_k = [\boldsymbol{Y}_k, \boldsymbol{Y}_{k-1}, \ldots, \boldsymbol{Y}_1, \boldsymbol{Y}_0]\,, \tag{C.12}$$

where $\boldsymbol{\theta}$ is a vector of unknown model parameters and the initial values, if these are to be estimated. Provided that the sample intervals are reasonable compared to the non-linear dynamics of the system, the conditional probability densities p can be approximated by Gaussian densities, as the SDE (C.1) is driven by a Wiener process which has Gaussian increments. With this assumption the probability density is

$$p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta}) = \frac{\exp(-\frac{1}{2}\boldsymbol{\epsilon}_k^T(\boldsymbol{\Sigma}_{k|k-1}^{yy})^{-1}\boldsymbol{\epsilon}_k)}{\sqrt{\det\left(\boldsymbol{\Sigma}_{k|k-1}^{yy}\right)}\left(\sqrt{2\pi}\right)^l}\,, \tag{C.13}$$

where

$$\hat{\boldsymbol{Y}}_{k|k-1} = \mathrm{E}[\boldsymbol{Y}|\mathscr{Y}_{k-1},\boldsymbol{\theta}]\,, \tag{C.14}$$

$$\boldsymbol{\Sigma}_{k|k-1}^{yy} = \mathrm{Var}[\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta}]\,, \text{ and} \tag{C.15}$$

$$\boldsymbol{\epsilon}_k = \boldsymbol{Y}_k - \hat{\boldsymbol{Y}}_{k|k-1}\,. \tag{C.16}$$

If prior information about all or some of the parameters is available, maximum a posteriori (MAP) estimation can be used. Assuming that the prior probability

density of the parameters is Gaussian, the likelihood function can be written

$$L(\boldsymbol{\theta};\mathscr{Y}_N) = \left(\prod_{k=1}^{N} p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta})\right) p(\boldsymbol{Y}_0|\boldsymbol{\theta}) \cdot \frac{\exp(-\frac{1}{2}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\epsilon}_{\boldsymbol{\theta}})}{\sqrt{\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})}\left(\sqrt{2\pi}\right)^p}, \tag{C.17}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = E[\boldsymbol{\theta}], \tag{C.18}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = V[\boldsymbol{\theta}], \tag{C.19}$$

$$\boldsymbol{\epsilon}_{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}, \tag{C.20}$$

and $p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta})$ is given in equation (D.25). The conditional probability density $p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta})$ is fully described by the mean and variance of the measurements, which can be estimated recursively by means of the Extended Kalman Filter, EKF. First some more notation is introduced

$$\boldsymbol{\Sigma}_{k|k-1}^{xx} = \mathrm{Var}[\boldsymbol{X}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta}], \tag{C.21}$$

and

$$\boldsymbol{A} = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{X}_t}\right|_{\boldsymbol{X}=\hat{X}_{k|k-1},\boldsymbol{u}=\boldsymbol{u}_k,t=t_k,\boldsymbol{\theta}}, \tag{C.22}$$

$$\boldsymbol{C}_k = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{X}_t}\right|_{\boldsymbol{X}=\hat{X}_{k|k-1},\boldsymbol{u}=\boldsymbol{u}_k,t=t_k,\boldsymbol{\theta}}, \tag{C.23}$$

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}), \text{ and} \tag{C.24}$$

$$\boldsymbol{R}_k = \boldsymbol{R}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}), \tag{C.25}$$

The EKF consists of two sets of equations, namely the updating and the prediction equations:
The updating equations are

$$\hat{\boldsymbol{X}}_{k|k} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\epsilon}_k, \text{ and} \tag{C.26}$$

$$\boldsymbol{\Sigma}_{k|k}^{xx} = \boldsymbol{\Sigma}_{k|k-1}^{xx} - \boldsymbol{K}_k\boldsymbol{\Sigma}_{k|k-1}^{yy}\boldsymbol{K}_k^T, \tag{C.27}$$

with the Kalman gain

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k-1}^{xx}\boldsymbol{C}_k^T\left(\boldsymbol{\Sigma}_{k|k-1}^{yy}\right)^{-1}. \tag{C.28}$$

The prediction equations are

$$\frac{d\hat{\boldsymbol{X}}_{t|k}}{dt} = \boldsymbol{f}(\hat{\boldsymbol{X}}_{t|k}, \boldsymbol{u}_t, t, \boldsymbol{\theta}), \text{ and} \tag{C.29}$$

$$\frac{d\boldsymbol{\Sigma}_{t|k}^{xx}}{dt} = \boldsymbol{A}\boldsymbol{\Sigma}_{t|k}^{xx} + \boldsymbol{\Sigma}_{t|k}^{xx}\boldsymbol{A}^T + \boldsymbol{\sigma}\boldsymbol{\sigma}^T, \tag{C.30}$$

which is solved for $t \in [t_k, t_{k+1}[$, and finally the prediction equations for the mean and covariance of the observations are given by

$$\hat{\boldsymbol{Y}}_{k|k-1} = \boldsymbol{h}(\hat{\boldsymbol{X}}_{t|k-1}, \boldsymbol{u}_t, t_k, \boldsymbol{\theta}) \text{ , and} \tag{C.31}$$

$$\boldsymbol{\Sigma}^{yy}_{k|k-1} = \boldsymbol{C}_k \boldsymbol{\Sigma}^{xx}_{k|k-1} \boldsymbol{C}_k^T + \boldsymbol{R} \,. \tag{C.32}$$

The initial values for the EKF are $\hat{\boldsymbol{X}}_{t|t_0} = \boldsymbol{X}_0$ and $\boldsymbol{\Sigma}^{xx}_{t|t_0} = \boldsymbol{\Sigma}^{xx}_0$, which can either be fixed or estimated along with the ML estimation.

Model validation - Step Ic

Different methods can be used to validate an SDE model. Goodness of fit can be evaluated by considering the standardized residuals r based on the one-step predictions $\hat{Y}_{k|k-1}$ (C.14). The predictions are generated by the extended Kalman filter. For the bacterial growth model the standardized residuals are given by

$$r_k = \frac{Y_k - \hat{Y}_{k|k-1}}{\Sigma^{yy}_{k|k-1}} \,. \tag{C.33}$$

According to the SDE model formulation the standardized residuals should be independent normally distributed. This can be checked e.g. by the autocorrelation function of the standardized residuals. We use the function `acf` in `R` (R Development Core Team, 2009) to compute the autocorrelation function. Additionally r can be plotted against e.g. time and the other state variables. If any of these plots do not correspond well with the model assumption, this indicates that there is something wrong with the model. For other analysis considering the residuals see e.g. (Kristensen et al., 2004b; Madsen, 2008).
Other tests for goodness of fit exist such as the one proposed by Bak et al. (1999), which uses simulation of trajectories between neighboring observations. However, going into details with this method is out of the scope of this article.
If the result of the validation is satisfactory, the best model has been found and no further steps need to be executed. If the validation is not satisfactory, there are two possibilities. If there are obvious ways of reducing (or extending) the model, then Part I should be repeated for the suggested model. Reduction of the model will often be related to test for significants of parameters or for reduction by setting two parameters equal. Next the models should be compared in the validation step. If the models are nested, a likelihood-ratio test can be used, which provides a p-value on which a decision can be based. Alternatively, Akaike's Information Criterion (AIC) given by

$$AIC = -2\log(L) + 2p \tag{C.34}$$

can be applied for non-nested models, where p is the number of parameters in the function. The model with the lowest AIC value should be selected.
If no obvious model changes are available Part II should be performed to identify improvements to the model.

C.3.2 Part II (non-parametric)

Non-parametric model extension - Step IIa

The part of the model that calls for improvement may be found e.g. by identifying the state variable with the highest diffusion coefficient. The rate parameter in the drift part of this state should then be included as an extra state depending only on the system noise. The rate is thus modelled by a non-parametric extension to the SDE.
Let us consider the model (C.9). Now we want to study the functional dependence of the growth rate on the substrate content. Therefore the growth rate is implemented as a new state variable, and the total system equation becomes

$$d\begin{bmatrix} B_t \\ S_t \\ \mu_t \end{bmatrix} = \begin{bmatrix} \mu_t B_t \\ -\eta\mu_t B_t \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sigma_B u_t & 0 & 0 \\ -\eta\sigma_B u_t & 0 & 0 \\ 0 & 0 & \sigma_\mu \end{bmatrix} d\boldsymbol{\omega}_t \,. \tag{C.35}$$

The observation equation is unchanged.

Parameter estimation - Step IIb

Corresponds to the method used for parameter estimation in Step 2, see Section C.3.1

Smoothing - Step IIc

The smoothing estimate is computed by a combination of a forward filter estimate based on a few past observations, and a backward filter estimate based on present and future observations. It can be shown (Gelb et al., 1974) that the optimal smoother $\hat{\boldsymbol{X}}_{k|N}$ is given by the following combination of the forward filtering estimate $\hat{\boldsymbol{X}}_{k|k-1}$ and backward filtering estimate $\overline{\boldsymbol{X}}_{k|k}$

$$\hat{\boldsymbol{X}}_{k|N} = \boldsymbol{\Sigma}^{xx}_{k|N}\left((\boldsymbol{\Sigma}^{xx}_{k|k-1})^{-1}\hat{\boldsymbol{X}}_{k|k-1} + \left(\overline{\boldsymbol{\Sigma}}^{xx}_{k|k}\right)^{-1}\overline{\boldsymbol{X}}_{k|k}\right), \tag{C.36}$$

where

$$\boldsymbol{\Sigma}^{xx}_{k|N} = \left(\left(\boldsymbol{\Sigma}^{xx}_{k|k-1}\right)^{-1} + \left(\overline{\boldsymbol{\Sigma}}^{xx}_{k|k}\right)^{-1}\right)^{-1}. \tag{C.37}$$

The forward filter estimate of the state variables, together with the covariance $\boldsymbol{\Sigma}^{xx}_{k|k-1}$, is determined by the EKF as described in equation (C.26)-(C.32). To determine the backward filter estimate $\overline{\boldsymbol{X}}_{k|k}$ and the corresponding covariance matrix $\overline{\boldsymbol{\Sigma}}^{xx}_{k|k}$ we need to first rewrite the state equation (C.1). As the backward filter starts at the end point and moves forward towards the smoothing point, we write the time as $\tau = t_N - t$. Thus, the state equation takes on the form

$$\begin{aligned} d\boldsymbol{X}_{t_N-\tau} = &-\boldsymbol{f}(\boldsymbol{X}_{t_N-\tau}, \boldsymbol{u}_{t_N-\tau}, t_N-\tau, \boldsymbol{\theta})d\tau \\ &-\boldsymbol{\sigma}(\boldsymbol{u}_{t_N-\tau}, t_N-\tau, \boldsymbol{\theta})d\boldsymbol{\omega}_\tau \,. \end{aligned} \tag{C.38}$$

In order to have well defined boundary conditions the following transformation is applied

$$\boldsymbol{s}_k = \left(\overline{\boldsymbol{\Sigma}}^{xx}_{k|k}\right)^{-1} \overline{\boldsymbol{X}}_{k|k} \,. \tag{C.39}$$

The smoothing estimate must equal the forward filter estimate at the end time N. This means that

$$\boldsymbol{\Sigma}^{xx}_{N|N} = \boldsymbol{\Sigma}^{xx}_{N|N-1} \,, \tag{C.40}$$

which according to equation (C.37) leads to

$$\left(\overline{\boldsymbol{\Sigma}}^{xx}_{N|N}\right)^{-1} = 0 \,, \tag{C.41}$$

whereby

$$\boldsymbol{s}_N = 0 \,. \tag{C.42}$$

Applying relation (C.39) the backward filter can be described by the following equations (Gelb et al., 1974):
The updating equations for $\boldsymbol{s}$ are

$$\boldsymbol{s}_{k|k} = \boldsymbol{s}_{k|k+1} + \boldsymbol{C}_k^T \boldsymbol{R}_k^{-1} \left(\boldsymbol{\epsilon}_k + \boldsymbol{C}_k \hat{\boldsymbol{X}}_{k|k-1}\right) \,, \tag{C.43}$$

$$\left(\overline{\boldsymbol{\Sigma}}^{xx}_{k|k}\right)^{-1} = \left(\overline{\boldsymbol{\Sigma}}^{xx}_{k|k+1}\right)^{-1} + \boldsymbol{C}_k^T \boldsymbol{R}_k^{-1} \boldsymbol{C}_k \,. \tag{C.44}$$

The prediction equations for $\boldsymbol{s}$ are

$$\begin{aligned} \frac{d\boldsymbol{s}_{t_N-\tau|k}}{d\tau} = \boldsymbol{A}_\tau^T \boldsymbol{s}_{t_N-\tau|k} - \left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1} \boldsymbol{\sigma}_\tau \boldsymbol{\sigma}_\tau^T \boldsymbol{s}_{t_N-\tau|k} \,, \\ - \left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1} \left(\boldsymbol{f}(\hat{\boldsymbol{X}}_{t|k}, \boldsymbol{u}_t, t, \boldsymbol{\theta}) - \boldsymbol{A}_\tau \hat{\boldsymbol{X}}_{t_N-\tau|k}\right) \end{aligned} \tag{C.45}$$

$$\begin{aligned} \frac{\left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1}}{dt} = \left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1} \boldsymbol{A}_\tau + \boldsymbol{A}_\tau^T \left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1} \\ - \left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1} \boldsymbol{\sigma}_\tau \boldsymbol{\sigma}_\tau^T \left(\overline{\boldsymbol{\Sigma}}^{xx}_{t_N-\tau|k}\right)^{-1} \,, \end{aligned} \tag{C.46}$$

which is solved for $t \in [\tau_k, \tau_{k+1}[$. In the equations for the backward filter the short hand notation

$$\boldsymbol{A}_\tau = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{X}_t}\right|_{\boldsymbol{X}=\hat{X}_{t_N-\tau|k}, \boldsymbol{u}=\boldsymbol{u}_{t_N-\tau}, t=t_N-\tau, \boldsymbol{\theta}} \tag{C.47}$$

$$\boldsymbol{\sigma}_\tau = \boldsymbol{\sigma}(\boldsymbol{u}_{t_N-\tau}, t_N-\tau, \boldsymbol{\theta}) \,, \tag{C.48}$$

is used, along with the notation for $\boldsymbol{C}_k$ and $\boldsymbol{R}_k$ given in (C.23) and (C.25).

C.3.3 Interpretation of smoothing results - step IId

The smoothing data generated as described in Section C.3.2 can be used to dig into embedded information of the dynamics of the system by plotting the relation between the new state variable, which in this case is the growth rate, and each of the other states as well as time. These plots, together with physical insight about the system, should then be used to improve the model. Either by suggesting a new expression relating the phenomenon of interest with one or more other states, or by introducing an extra state to the system. This new suggestion for a candidate model should hereafter be tested by the steps in Part I.

C.4 Simulation study

A simulation study is carried out to demonstrate the steps in Part II on a growth curve with a known growth kinetics expression. The growth kinetics description used is the Monod equation (C.8). The set of SDEs describing the bacterial growth in Equation (C.9), where u_t is replaced by the state B_t, is solved numerically using an Euler approach implemented in `Matlab` with the parameters given in Figure C.2. Simulations are made in triplicates for three different initial optical density (OD) values: 0.05, 0.005 and 0.0005. To resemble observation noise, normal distributed noise is added to the simulated data, resulting in the time series shown in Figure C.2. The parameters of the reformulated model (C.35) are estimated for

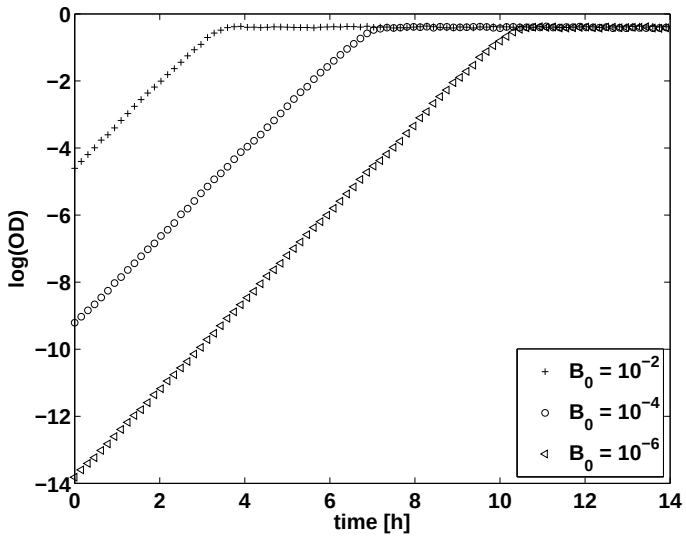


Figure C.2: Change of bacterial concentration over time as simulated with the SDE (C.9) and Monod growth kinetics (C.8) for three different initial optical density (OD) values. Normal distributed noise with a standard deviation of 0.001 is added to the simulated data to resemble observation noise. The parameter values used for the simulation are: $\nu = 1.6h^{-1}$, $\kappa = 0.2$, $\eta = 1.5$, and $\sigma_B = 0.05$.

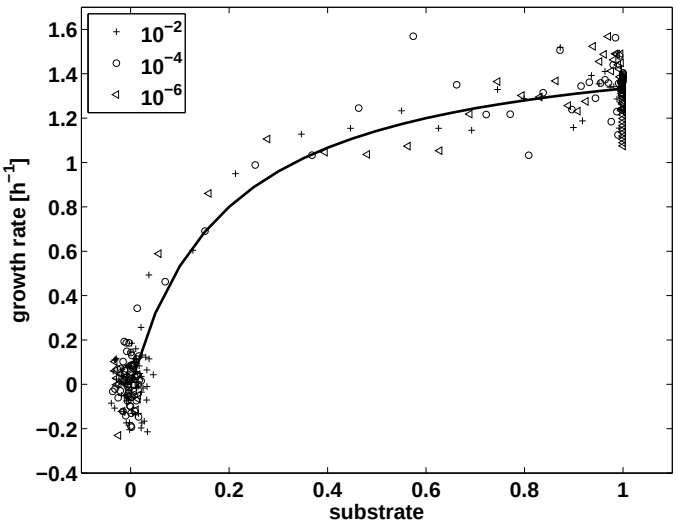


Figure C.3: Smoothing results for one repetition of each initial concentration. The smoothing data seem to capture well the Monod shape of the growth rate used for the simulation, which is shown with the black line.

the constructed time series, and smoothing data are created. The result for one repetition of each initial concentration is shown in Figure C.3, where the result of the smoothing is plotted together with the known Monod expression for the growth kinetics. The method is seen to reconstruct very well the relation between the growth rate and the substrate level.

C.5 Data

Optical density measurements for the growth of *Salmonella* (S67) and *Enterococcus* (E46) growing in BHI media are available for the analysis. For each bacteria strain a 9-fold dilution as well as the non-diluted strain are measured in duplicates. The measurements are made for 40 hours with a sampling interval of 20 minutes in a bioscreen (Microbiology Reader Bioscreen C) at 16°C under continuous shaking.
The OD measurements are preliminarily corrected for background broth measurements and subsequently a correction is made for the shadow effect for high OD values by the relation (Philipsen et al., 2010b)

$$\mathrm{OD}_{\mathrm{meas}} = a(1 - \exp(-b \cdot \mathrm{OD}_{\mathrm{corr}})). \tag{C.49}$$

The constants a and b are found by fitting the relation to results from an experiment with *Salmonella* for which the OD for different concentrations has been measured. It is assumed that the same parameters can be used for *Enterococcus*. The OD measurements after corrections are shown in Figure C.4.

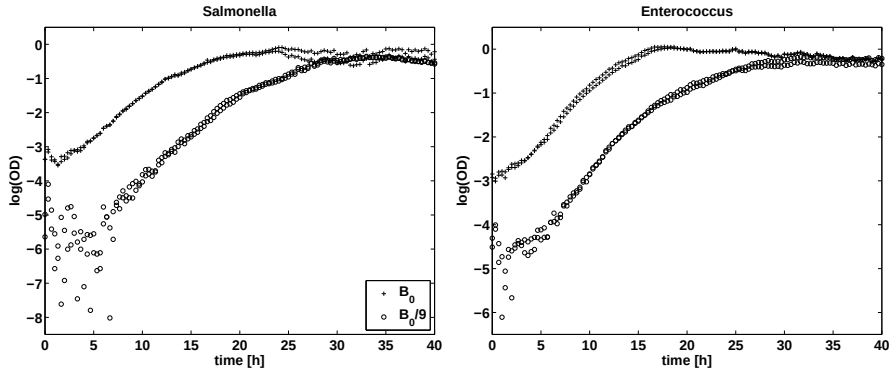


Figure C.4: Growth curves for the four different bacterial populations used in the experiment. The background OD has been subtracted from the measured OD and the OD has been corrected for the shadow effect at higher OD values as given by (C.49).

From the observations we obtain estimates of η calculated from equation (C.7),

$$\begin{aligned} \textit{Salmonella}: \quad & E[\eta] = 1.34\,, \quad V[\eta] = 0.13 \\ \textit{Enterococcus}: \quad & E[\eta] = 1.34\,, \quad V[\eta] = 0.11 \end{aligned}$$

This prior information about η can be used to perform a MAP estimate of this parameter, when estimating all the parameters of the SDE model in Equation (C.9) or Equation (C.35).

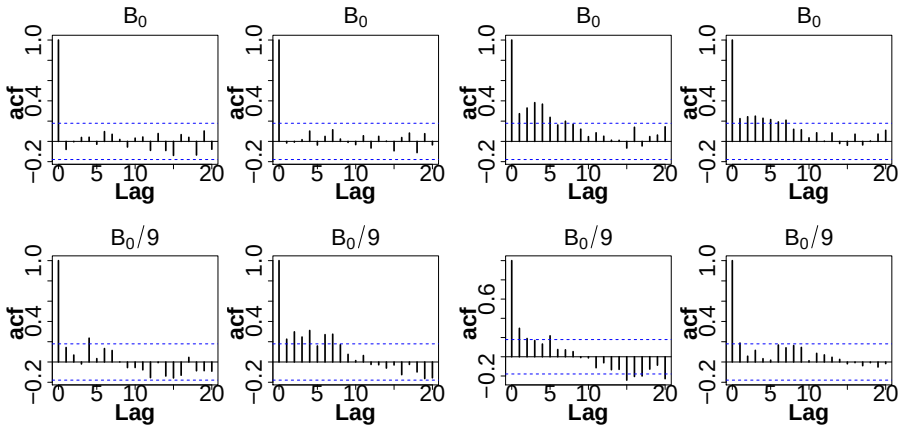
C.6 Results

C.6.1 Part I

The parameters for the SDE in Equation (C.9) using the Monod relation Equation (C.8) have been estimated and one-step predictions have been generated using `CTSM`. The model is validated by considering the standardized residuals. From the autocorrelation function plotted in Figure C.5 it is seen that the standardized residuals are not uncorrelated for all data sets. In several cases there is autocorrelation for lags 1 to 5. As there are no obvious model reductions we continue to Part II.

C.6.2 Part II

The model has been reformulated as written in Equation (C.35) and the parameters are estimated. The results of the smoothing for the two different bacterial growth curves can be seen in Figure C.6. The challenge is now to suggest a candidate model based on the result of the smoothing. By comparing Figure C.6 (bottom) and Figure C.4 it is seen that the growth rate increases until after the end of the lag phase for the growth, during which the substrate level stays constant as



(a) *Salmonella* (b) *Enterococcus*

Figure C.5: Autocorrelation function for the standardized residuals from the Monod growth.

seen in Figure C.6 (top). Due to these time-dependent dynamics of the growth rate, it is necessary to include an extra state in the system equation. One approach is to use the optimal model which has been introduced to model growth on multiple substrates (Bajpai-Dikshit et al., 2003). This model is based on the assumption that one enzyme is rate-limiting for catalyzing the consumption of substrate. The dynamics of the enzyme level E is typically described as

$$\frac{dE_t}{dt} = \frac{(v+\beta)S_t}{\kappa + S_t} - E_t \frac{d\log(B_t)}{dt} - \beta E_t\,, \tag{C.50}$$

for which the Monod expression has traditionally been used to model growth kinetics (Doshi et al., 1997; Kompala et al., 1984; Bajpai-Dikshit et al., 2003). In the optimal model, two states for substrate and enzyme levels, respectively, are included in the model per substrate in the experimental solution. However, the smoothing data do not indicate that more than one state for the substrate is necessary.

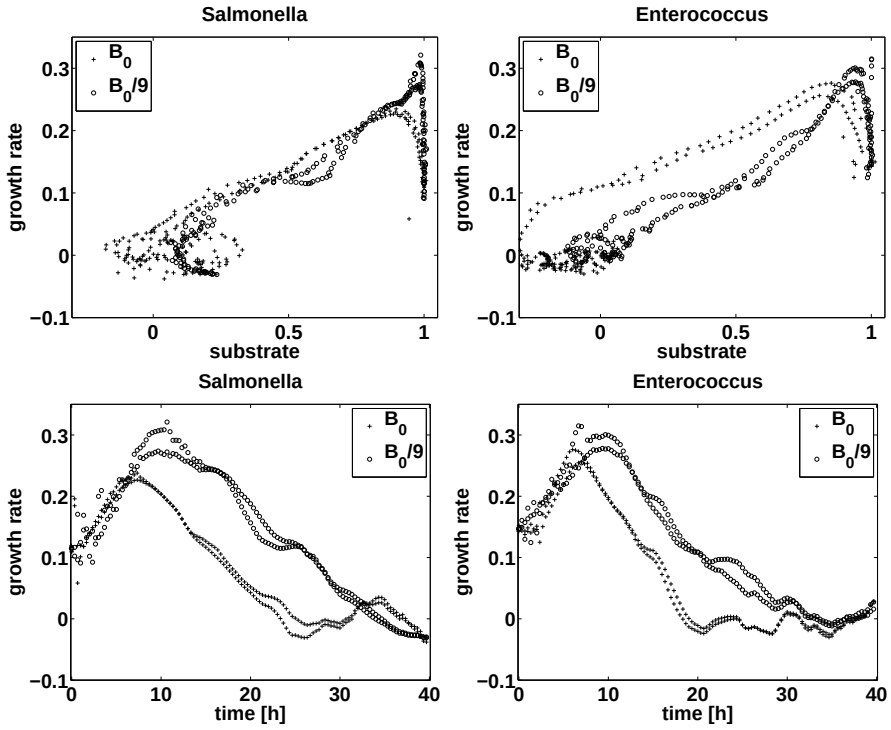


Figure C.6: Smoothing data created from Equation (C.35). The growth rate as a function of substrate (top) and time (bottom) is plotted.

C.6.3 Part I

Adding an extra state for the enzyme level to the system equation, and using the Monod growth equation, the new candidate model is

$$d\begin{bmatrix} B_t \\ S_t \\ E_t \end{bmatrix} = \begin{bmatrix} E_t \upsilon \lambda_t B_t \\ -\eta E_t \upsilon \lambda_t B_t \\ (\upsilon+\beta)\lambda_t - E_t E_t \upsilon \lambda_t - \beta E_t \end{bmatrix} dt + \begin{bmatrix} \sigma_B u_t & 0 & 0 \\ -\eta \sigma_B u_t & 0 & 0 \\ 0 & 0 & \sigma_E \end{bmatrix} d\omega\,. \quad \text{(C.51)}$$

where

$$\lambda_t = \lambda_0 = \frac{S_t}{\kappa + S_t} \quad \text{(C.52)}$$

The estimated parameters for this model are given in the first two columns in Table C.1 and the resulting simulation is compared with data in Figure C.7. The estimates of υ and κ are strongly correlated; 0.9998 and 0.9923 for *Salmonella* and *Enterococcus* respectively. This indicates that the model is over-parameterized.

Table C.1: Parameter values (standard deviation) for the growth model with λ_0 and λ_1

	λ_0		λ_1	
	Salmonella	*Enterococcus*	*Salmonella*	*Enterococcus*
κ	12.28 (32.56)	95.57 (37.62)	-	-
υ	3.648 (8.969)	31.20 (12.40)	0.272 (0.010)	0.281 (0.007)
β	0.011 (0.021)	0.093 (0.038)	0.006 (0.020)	0.091 (0.043)
η	1.429 (0.051)	1.134 (0.030)	1.420 (0.051)	1.134 (0.037)
σ_B	0.073 (0.003)	0.042 (0.001)	0.073 (0.003)	0.042 (0.002)
σ_E	0.0001 ($1{\cdot}10^{-5}$)	0.0001 ($3{\cdot}10^{-5}$)	0.0001 ($6{\cdot}10^{-6}$)	0.0001 ($9{\cdot}10^{-6}$)
$\sqrt{R}$	0.0026 (0.0002)	0.0027 (0.0002)	0.0026 (0.0002)	0.0027 (0.0002)
B_0	0.035 (0.001)	0.056 (0.001)	0.035 (0.001)	0.056 (0.001)

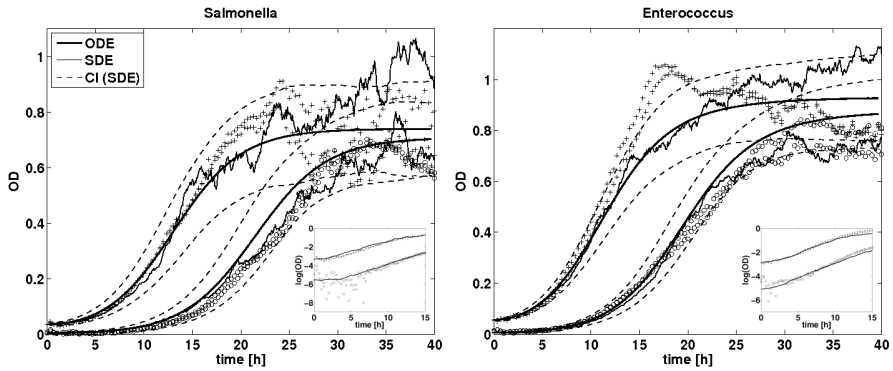


Figure C.7: Simulation of the growth model in Equation C.51 with a growth rate proportional to the substrate level. One realization of the SDE model is shown together with the confidence interval (CI). The SDE simulation is computed in `Matlab` with an implementation of the Euler method for numerical solution of SDEs.

Table C.2: AIC values for the different growth rate expressions

	λ_0	λ_1
Salmonella	-2884.3	-2886.0
Enterococcus	-5868.8	-5872.6

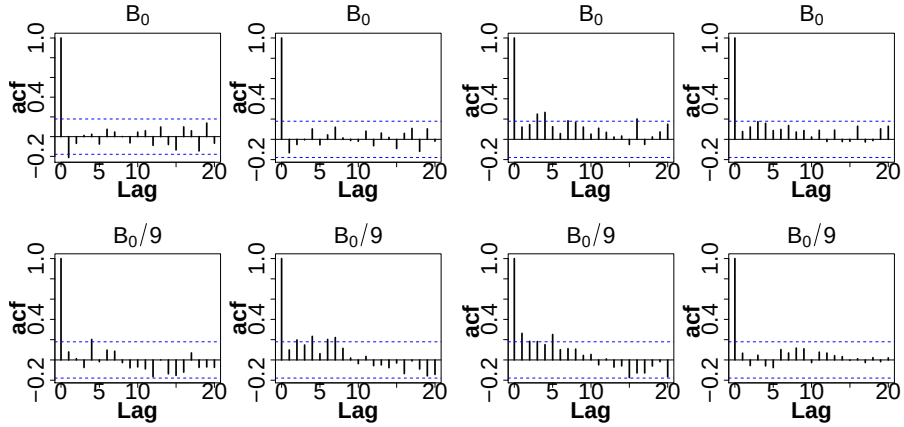
Additionally the estimated κ values are high (12.277 and 95.570) compared to the maximum value of the substrate content on 1. Hereby the substrate dependence becomes approximately proportional. Thus, an obvious model reduction is

$$\lambda_t = \lambda_1 = S_t \tag{C.53}$$

The parameters for this reduced model are estimated and the result is given in the last two columns of Table C.1. The models with λ_0 and λ_1 are tested against each other using AIC. The result of the test is given in Table C.2, and the best model is

found to be with λ_1, i.e. a model with a growth rate proportional to the substrate content.

In order to further validate the model with the extra state for the enzyme level



(a) *Salmonella* (b) *Enterococcus*

Figure C.8: Standardized residuals for the model Equation (C.51) with $\lambda_1 = S$.

and with λ_1, the autocorrelation function of the standardized residuals is plotted in Figure C.8. An improvement has been made to the autocorrelation function in comparison to Figure C.5, as now only few lags have significant autocorrelations, and the size of the autocorrelation has generally been reduced. Whether the standardized residuals can be considered as white noise can be tested using Box-Pierce test statistics (Madsen, 2008). For *Salmonella* it is found that the p-value increases from the initial model to the new model from 0.0027 to 0.11. This means that for the new model the hypothesis of independent standardized residuals for *Salmonella* can not be rejected. For *Enterococcus* the hypothesis of white noise is rejected with a p-value of 0.003, but the test statistic has been reduced by approximately half compared to the first model, which indicates a strong improvement of the model. Further improvements of the model would demand additional data or information about the growth process, but for now we are satisfied with the model.

C.7 Conclusion

The systematic framework for model improvement presented in this study is a strong tool that combines physical knowledge about a system with statistical methods to obtain the best model. The framework is built on a modelling technique based on stochastic differential equations, which facilitates identification of model deficiencies.

We have found that the growth of bacteria on rich media can be adequately modelled by including one extra state representing the rate-limiting enzyme level which

accounts for the initial lag phase of the growth. The growth rate for bacteria growing on minimal media is traditionally modelled using the Monod equation, but we find that a growth rate proportional to the substrate level gives a better fit to the data.

Acknowledgements

This study was supported by the Danish Research Council for Technology and Production Sciences through the grant 274-05-0117 titled "Evolution and adaptation of antimicrobial resistance in bacterial populations".

C.8 Bibliography

Bajpai-Dikshit, J., Suresh, A. K., Venkatesh, K. V., 2003. An optimal model for representing the kinetics of growth and product formation by lactobacillus rhamnosus on multiple substrates. Journal of Bioscience and Bioengineering 96 (5), 481–486.

Bak, J., Nielsen, H. A., Madsen, H., 1999. Goodness of fit of stochastic differential equations. In: Symposium i Anvendt Statistik. pp. 341–346.

Doshi, P., Rengaswamy, R., Venkatesh, K. V., 1997. Modelling of microbial growth for sequential utilization in a multisubstrate environment. Process Biochemistry 32 (8), 643–650.

Gelb, A., Kasper Jr., J. F., Nash Jr., R. A., Sutherland Jr., A. A., 1974. Applied Optimal Estimation. The M.I.T. press.

Kompala, D. S., Ramkrishna, D., Tsao, G. T., 1984. Cybernetic modeling of microbial growth on multiple substrates. Biotechnology and Bioengineering 26 (11), 1272–1281.

Kovárová-Kovar, K., Egli, T., 1998. Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics. Microbiology and Molecular Biology Reviews 62 (3), 646–666.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004a. A method for systematic improvement of stochastic grey-box models. Computers and Chemical Engineering 28, 1431–1449.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004b. Parameter estimation in stochastic grey-box models. Automatica 40 (40), 225–237.

Øksendal, B., 2007. Stochastic Differential Equations. Springer.

Liu, Y., 2006. A simple thermodynamic approach for derivation of a gereral monod equation for microbial growth. Biochemical Engineering Journal 31, 102–105.

Lobry, J., Flandrois, J., Carret, G., Pave, A., 1992. Monod's bacterial growth model revisited. Bulletin of Mathmatical Biology 54, 117–122.

Madsen, H., 2008. Time series analysis. Chapman & Hall/CRC.

Monod, J., 1949. The growth of bacterial cultures. Annual Review of Microbiology 3, 371–394.

Patnaik, P. R., 1999. Transient sensitivity analysis of a cybernetic model of microbial growth on two substrates. Bioprocess Engineering 21 (2), 135–140.

Philipsen, K., Christiansen, L., Hasman, H., Madsen, H., 2010a. Modelling conjugation with stochastic differential equations. Journal of Theoretical Biology 263 (1), 134–142.

Philipsen, K. R., Christensen, L. E., Mandsberg, L. F., Hasman, H., Madsen, H., 2010b. Comparison of calibration curves for the relation between optical density and viable count bacteria data, submitted.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL `http://www.R-project.org`

PAPER

D

Modelling conjugation with stochastic differential equations [‡]

Abstract

Conjugation is an important mechanism involved in the transfer of resistance between bacteria. In this article a stochastic differential equation based model consisting of a continuous time state equation and a discrete time measurement equation is introduced to model growth and conjugation of two *Enterococcus faecium* strains in a rich exhaustible media. The model contains a new expression for a substrate dependent conjugation rate. A maximum likelihood based method is used to estimate the model parameters. Different models including different noise structure for the system and observations are compared using a likelihood-ratio test and Akaike's information criterion. Experiments indicating conjugation on the agar plates selecting for transconjugants motivates the introduction of an extended model, for which conjugation on the agar plate is described in the measurement equation. This model is compared to the model without plate conjugation. The modelling approach described in this article can be applied generally when modelling dynamical systems.

D.1 Introduction

Development and spread of antimicrobial resistance in bacterial populations is of increasing concern, as it can lead to major difficulties for the treatment of diseases. A first step in the direction of solving this problem is to gain a better understanding of the spread of resistance, and here conjugation plays a major role. Conjugation is one of several mechanisms of horizontal gene transfer by which plasmids coding for e.g. antimicrobial resistance can be transferred between bacteria. The use of mathematical modelling to describe the dynamics of plasmid spread and

[‡]As published in: K. R. Philipsen, L. E. Christiansen, H. Hasman, H. Madsen, 2009. Modelling conjugation with stochastic differential equations. Journal of Theoretical Biology 263 (1), 134-142

persistence was first introduced by Levin et al. (1979) and since then many studies have been made to improve the model framework and parameter estimation and to incorporate more accurate expressions for the plasmid dynamics (Slater et al., 2008). Most studies have used ordinary differential equations, ODE (Levin et al., 1979; Freter et al., 1983; Clewlow et al., 1990; MacDonald et al., 1992; Willms et al., 2006), or ordinary difference equations (Knudsen et al., 1988; Sudarshana and Knudsen, 2006) as the modelling framework. This approach can be discussed since a part of the complexity of conjugation, e.g. dependence on the surrounding environment, is not included in these models. A way of overcoming this problem is to use a stochastic modelling approach where the randomness accounts for those processes not included in the model. Some efforts have been made to include randomness in plasmid models (Joyce et al., 2005; De Gelder et al., 2007; Ponciano et al., 2007). Common for these studies is that they use discrete-time models, which is a good approximation at low bacteria densities, but not at higher densities where growth and conjugation is a continuous process. Joyce et al. (2005) give a first and second order moment representation of their model which requires a derivation of the means and variances. Thus, it is not easy to transfer the method to other models. In the study by Ponciano et al. (2007) one of the model parameters is implemented as a normal distributed random variable. This paper presents a generic modelling framework first described by Kristensen et al. (2004a), which is based on stochastic differential equations, SDE. The SDE is connected to data through a state-space formulation consisting of continuous-time state equations (the SDE) and discrete-time observation equations. The applied state-space approach is in fact a continuous time hidden Markov model. The state-space formulation opens up for strong statistical tools for estimating model parameters and for inference concerning the best model (Kristensen et al., 2004a). State-space modelling has also been used in previous plasmid studies (De Gelder et al., 2004, 2007; Ponciano et al., 2007), but our approach differs by enabling continuous-time state equations in combination with discrete-time observations. The state-space model enables a simultaneous estimation of the growth and conjugation parameters, whereby the bacterial growth is accounted for when estimating the conjugation rate. This is an improvement from previous studies, where the donor and recipient concentrations were introduced as a mean value of two measurements taken over time (Knudsen et al., 1988; Sudarshana and Knudsen, 2006) or the experiment was constructed such that bacterial growth could be neglected (MacDonald et al., 1992).

In this article the SDE based state-space modelling framework is applied to analyze data from an *in vitro* experiment for conjugation between two *Enterococcus faecium* species growing in a batch exhaustable media. This experiment was made in order to study the transferability of vancomycin resistance in *E. faecium*. To describe the experimental system a conjugation model of bacteria growing in a broth exhaustible media is introduced, which is an expansion of previous models. Several authors (Levin and Stewart, 1980; Freter et al., 1983; Knudsen et al., 1988; Clewlow et al., 1990; Top et al., 1992; Willms et al., 2006; Sudarshana and Knudsen,

2006) have modelled conjugation events with the mass action model proposed by Levin et al. (1979). The mass action model states that the appearance of transconjugants is proportional to the product of the donor and recipient concentrations. We will introduce a new expression for the conjugation rate in an exhaustible media for which the proportionality constant of the mass action model is substrate dependent. An inference study is made to reduce the SDE model to its minimum form. A further extension of the model is made to treat adequately a methodological problem: the finding that conjugation continues on the agar plates selecting for transconjugants. In that case the observed transconjugants are a combination of transconjugants stemming from the conjugation process in the flask and conjugation occurring on the selective plates. The models with and without conjugation on the selective plates are compared in order to examine which model best describes data.

D.2 Experimental methods

A conjugation experiment was made in order to study the transferability of vancomycin resistance in *Enterococcus faecium*. As recipient the *E. faecium* reference strain BM4105RF was used. This strain is resistant to rifampicin (MIC > 25 μg/ml) and fusidic acid (MIC > 25 μg/ml) due to chromosomally located mutations. As donor, the *E. faecium* A17sv1 (Hasman and Aarestrup, 2002) was used. This strain is resistant to erythromycin (MIC > 16 μg/ml) and vancomycin (MIC > 32 μg/ml) due to the presence of the *erm*(B) gene and *Tn*1546 transposon (carrying the *vanA*-gene cluster), respectively, located on a conjugative plasmid.

The conjugation experiment was performed in liquid Brain-Heart Infusion (BHI) media (Oxoid). Bacterial counting was performed on BHI agar plates supplemented with the following antibiotics when appropriate: rifampicin 25 μg/ml, fusidic acid (25 μg/ml), erythromycin (16 μg/ml) and vancomycin (32 μg/ml).

D.2.1 Conjugation experiment

Over night blood agar cultures of the two strains grown at 37°C were inoculated in BHI media supplemented with the appropriate antibiotics as described above. From these, 100 μl of each culture was transferred to fresh tubes containing 10 ml preheated BHI media without antibiotics supplementation. When these cultures reached late exponential growth (OD_{500} of 0.3-0.5), the number of cells in each culture was adjusted to the same amount and 1.5 ml of each was transferred to 100 ml preheated (37°C) BHI media in an 250 ml Erlenmeyer flask. The flask was placed in a shaking incubator (125 rpm) at 37°C. Four 1 ml samples were taken immediately after the cells were added (t=0) as well as 1.50, 2.80, 3.00, 3.40, 3.80, 4.00, 4.40, 4.80, 5.00, 5.50, 6.00, 6.30 and 7.00 hours after the cells were mixed. One of the four 1 ml samples was used to measure the OD_{500} of the culture. The remaining three samples were diluted 10-fold until 10^{-7}. From here, 100 μl of each dilution (where appropriate) was plated onto sets of plates containing BHI

agar supplemented with either of the following: 1) recipient plates containing 25 μg/ml rifampicin + 25 μg/ml fusidic acid, 2) Donor plates containing 16 μg/ml erythromycin + 32 μg/ml vancomycin, and 3) transconjugant plates containing 25 μg/ml rifampicin + 25 μg/ml fusidic acid +16 μg/ml erythromycin + 32 μg/ml vancomycin. All plates were incubated 48 hours and the colony forming units, CFU, were counted. From these numbers the bacterial concentration in the flask was calculated.

D.2.2 Calculating bacterial concentration

The concentrations of the different bacteria populations in the flask can be determined from the plate count by a generalized linear model approach, assuming that the CFU count for each plate is Poisson distributed with an offset corresponding to the given dilution. If, for any given observation time t_k, N_{jk} is the count number j with dilution n_j, then the expected CFU count $E(N_{jk})$ can be modelled with the generalized linear model

$$E(N_{jk}) = \lambda_{jk} = n_j \exp(\beta_k)\,,\ N_{jk} \sim \text{Poisson}(\lambda_{jk}) \tag{D.1}$$

$$\log(\lambda_{jk}) = \log(n_j) + \beta_k\,. \tag{D.2}$$

Fitting the model to data gives an estimate for the coefficient β_k and hereby an estimate of the bacterial concentration $Y_k = \exp(\beta_k)$ in the flask at time, t_k. The variance $\sigma^2_{Y,k}$ of the estimated concentration at time t_k can be determined from the variance of β as

$$\sigma^2_{Y,k} = \text{Var}[\exp(\beta_k)] = (\exp(\beta_k)\text{Var}[\beta])^2\,. \tag{D.3}$$

The concentrations Y_k calculated in this way are the observations to be used for the modelling procedure described in the reminder of this article. The command `glmfit` in `Matlab` is used for fitting the generalized linear model (see Appendix D.A).

D.3 Model formulation

The modelling framework used in this study is a continuous-discrete time state-space model consisting of a continuous time state equation (the SDE) and a discrete time observation equation. The model has the general form (Kristensen et al., 2004b)

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t\,, \tag{D.4}$$

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k\,, \tag{D.5}$$

where (D.4) is the SDE and (D.5) is the observation equation. $\boldsymbol{X}_t$ is a n-dimensional vector of state variables and $\boldsymbol{Y}_k$ is a l-dimensional vector of observations. The observations are obtained at discrete times t_k with the observation noise being

$\boldsymbol{e}_k \in N(\mathbf{0}, \boldsymbol{\Sigma}_k)$. $\boldsymbol{\theta}$ is a vector of unknown parameters and $\boldsymbol{u}$ is a vector of input variables, i.e. variables which can be observed and have an influence on the system dynamics. The functions $\boldsymbol{f}$ and $\boldsymbol{h}$ can be linear as well as nonlinear functions. $\{\boldsymbol{\omega}_t\}$ is a standard Wiener process representing sources of noise in the system. The first part on the right-hand side of the SDE (D.4) is called the drift term and the second part is called the diffusion term. The stochastic state-space model (D.4)-(D.5) has several advantages compared to deterministic models. For instance, the state-space approach separates the residual noise into system noise and observation noise, where the system noise is used to e.g. compensate for those biological processes not explicitly described by the model. The deterministic model often leads to autocorrelated residuals, which is not a problem with the SDE. Using the SDE based model also paves the way for strong statistical tools to estimate model parameters and make inferences.

D.3.1 The drift term - Model for conjugation

In this study we apply the state-space modelling framework on a model for conjugation. The first step of the modelling procedure is to formulate the drift part of the SDE based on microbiological knowledge of the system. A sketch of the conjugation dynamics is given by the flow diagram in Figure D.1, where D is the donor, R the recipient, and T the transconjugant concentrations. The flow diagram shows

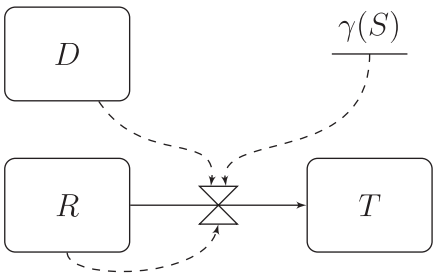


Figure D.1: Flow diagram showing the structure of the conjugation models. The donor D can by encounter of a recipient R transfer a plasmid coding for resistance to the recipient. The recipient hereby is turned into a transconjugant T, which expresses the same antibiotic resistances as both the donor and the recipient.

how recipients can be turned into transconjugants with a rate dependent on the concentration of donor and recipient. In earlier studies (MacDonald et al., 1992; Knudsen et al., 1988; Sudarshana and Knudsen, 2006) the substrate concentration and/or donor and recipient concentrations were kept constant in the experiments and were thus held constant in the model. In order to better approximate *in vivo* situations, these constraints were not applied to our *E. faecium* conjugation experiment, and thus all three populations as well as the substrate content must be included as state variables in the model. The drift term corresponding to the

model shown in Figure D.1 is

$$\boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{\theta}) = \begin{bmatrix} \mu_D(S)D \\ \mu_R(S)R - \gamma(S)DR \\ \mu_T(S)T + \gamma(S)DR \\ -\eta(\mu_D(S)D + \mu_R(S)R + \mu_T(S)T) \end{bmatrix} \tag{D.6}$$

where $\mu_i(S)$ is the growth rate for the bacterial population i (D, R or T), and $\gamma(S)$ is the conjugation rate. $\boldsymbol{X} = [D, R, T, S]$ are the model state variables and $\boldsymbol{\theta} = [\upsilon_D, \upsilon_R, \upsilon_T, \kappa_D, \kappa_R, \kappa_T, \gamma_{\max}, \kappa_c, \eta]$ are the parameters for the model. The substrate, S, is simulated as a normalized variable with an initial value of 1. The amount of substrate in the solution decreases as it is utilized for bacterial growth. The parameter η is the amount of normalized substrate used for each cell division. Bacterial growth continues until the substrate is exhausted, and the stationary state has been reached. The growth rate can be modelled with the well known Monod relation (Monod, 1949)

$$\mu_i(S) = \frac{\upsilon_i S}{\kappa_i + S}\,, S \in [0, 1]\,, \tag{D.7}$$

where S is the substrate concentration, υ_i is the maximum growth rate and κ_i is the substrate concentration when the growth rate is half of its maximum value. Both parameters are specific for a given bacterial population i.
The conjugation event depends on the probability of an encounter between donor and recipient. In the original mass action model the conjugation rate γ was introduced as a constant parameter. Several authors (Knudsen et al., 1988; Andrup et al., 1998; Andrup and Andersen, 1999; Ponciano et al., 2007) have discussed this assumption. Levin et al. (1979) showed that the mass action model presents a good estimation of the transconjugant population during exponential growth and under chemostatic conditions, but the model fails to describe the occurrences of transconjugants during the lag phase and at the onset of stationary phase. MacDonald et al. (1992) suggested that the conjugation rate depends on the substrate content and stated that conjugation can not occur without the presence of nutrition. However, to our knowledge this substrate dependence has not before been implemented in a mathematical model for conjugation. We suggest a nonlinear substrate dependent expression to model the conjugation rate

$$\gamma(S) = \frac{\gamma_{\max} S}{\kappa_c + S}\,, S \in [0, 1]. \tag{D.8}$$

This expression is similar to the Michaelis-Menten equation for enzyme kinetics and to the Monod relation. This expression is chosen as it forces the conjugation rate to reach a maximum value $\gamma_{\max}$ when the substrate concentration is abundant, and it turns to zero as the substrate is depleted. Depending on the value of κ_c the conjugation rate will decrease concurrent with or after the decrease in the bacterial growth rate.

Some assumptions are applied to keep the conjugation model simple. (i) Transconjugants can function as donors transferring a plasmid to a recipient, but this is not described by the model. It is assumed reasonable to omit it for the *E. faecium* data set used in this study, as the concentration of transconjugants is very low compared to the donor concentration, and therefore does not contribute significantly to conjugation. (ii) The delay on 10-15 min that has been found (Andrup et al., 1998) between two conjugation events for the same donor is assumed to be insignificant. During the experiment there is always a large number of donors not involved in a conjugation event and thus ready to start the conjugation by encounter of a recipient. Therefore the delay is disregarded. (iii) It is assumed that the maximum growth rate of the transconjugants is either the same as the maximum growth rate for the recipient or smaller due to a fitness cost of the plasmid, i.e. $\upsilon_T = \upsilon_R(1-\alpha)$, where $\alpha \in [0,1[$. An inference study will be made to test if $\alpha = 0$.

D.3.2 The diffusion term

Depending on the system which is modelled it can be adequate to implement additive system noise, i.e. noise independent on the state variables, or multiplicative system noise, i.e. where the noise depends on the state variables. The choice of the system noise depends on assumption about the system modelled. Tier and Floyd (1981) have described how different assumption of a biological process can lead to either demographic stochasticity (where the variance is proportional to the state variable) or environmental stochasticity (where the variance is proportional to the state variable squared). In our system the noise is implemented as multiplicative (environmental) noise, as the random fluctuations affect the whole population and not only the growth process. For instance the mass action model has shown to be good during exponential growth but not during lag-phase and stationary phase. Therefore multiplicative noise can be implemented to account for those processes not well described by the model. The method used for evaluating the likelihood function (Kristensen et al., 2004b) requires that the diffusion term is independent of the state variables. Therefore, instead of the state variables D, R and T, the input vectors $\boldsymbol{u}_D$, $\boldsymbol{u}_R$ and $\boldsymbol{u}_T$, which contains observations for the three states, are inserted as a scaling for the standard deviation. The noise term is

$$\boldsymbol{\sigma}_c^m = \begin{bmatrix} \sigma_D \boldsymbol{u}_D & 0 & 0 & 0 \\ 0 & \sigma_R \boldsymbol{u}_R & 0 & 0 \\ 0 & 0 & \sigma_T \boldsymbol{u}_T & 0 \\ -\eta\sigma_D \boldsymbol{u}_D & -\eta\sigma_R \boldsymbol{u}_R & -\eta\sigma_T \boldsymbol{u}_T & \sigma_S \end{bmatrix}, \tag{D.9}$$

where an increment in donor, recipient or transconjugant concentration lower the substrate content in order to keep the mass balance, and σ_S is introduced in order to ensure stability and is fixed to a small value.
An additive noise term $\boldsymbol{\sigma}_c^a$ is also introduced to compare with the multiplicative noise. This is done, as it is not sure that the data set is sufficient to make a good prediction of the system noise, and therefore a simplification of the noise structure

can be an advantage. The additive noise term is

$$\boldsymbol{\sigma}_c^a = \begin{bmatrix} \sigma_D & 0 & 0 & 0 \\ 0 & \sigma_R & 0 & 0 \\ 0 & 0 & \sigma_T & 0 \\ -\eta\sigma_D & -\eta\sigma_R & -\eta\sigma_T & \sigma_S \end{bmatrix} . \tag{D.10}$$

D.3.3 Observation equation

The observation equation relates observations to the state variables. In that way the state variables do not need to be measured directly and it is not necessary to have observations related to all state variables. In this study three of the state variables (donor, recipient and transconjugant) are observed. The concentration for the donor, recipient and transconjugant has been calculated from the CFU count as described in Section D.2.2. We will refer to this as the observed concentration. The observation equation, when only conjugation in the broth media is considered, is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}_k = \begin{bmatrix} D \\ R \\ T \end{bmatrix}_k + \boldsymbol{e}_k \,,\ \boldsymbol{e}_k \in N(\mathbf{0}, \boldsymbol{\Sigma}_k) \,. \tag{D.11}$$

The model with this observation equation is called the *Broth model*.
In addition to plasmid transfer in the broth, we have found that conjugation can occur on the transconjugant plates. This is a surprising result as antibiotics on the agar plate are traditionally considered to stop the conjugation process. However, we have performed several experiments (data not shown), which have revealed that conjugation does also occur on the transconjugant plates. As the experiment is made to measure conjugation in the broth media we wish to separate conjugation on the plate from that in the flask. The following set of ODEs is suggested to model conjugation on the plate for each observation time k

$$\frac{dD_{pk}}{dt} = -\lambda D_{pk} \,, \tag{D.12}$$

$$\frac{dR_{pk}}{dt} = -\lambda R_{pk} - \gamma_p D_{pk} R_{pk} \,, \tag{D.13}$$

$$\frac{dT_{pk}}{dt} = \gamma_p D_{pk} R_{pk} \,. \tag{D.14}$$

The initial values for the concentration of donor, D_{pk}, and recipient, R_{pk}, on the plate correspond to the concentrations in the flask at the time for the observation. Due to antibiotics on the plate, the donors and recipients die with a death rate λ and even though substrate is present in abundant amounts it is assumed that the donor and recipient bacteria can not grow. The recipients which receives a plasmid on the agar plate (and thus become a transconjugants) become resistant to all the antibiotics present and can therefore grow and form colonies on the plate. The

plate conjugation rate γ_p is assumed to be independent off the substrate content as the substrate will not be depleted before the conjugation process stops.
The plate conjugation is modelled by finding the analytical solution for the ODE and inserting the solution in the observation equation for transconjugants equation (D.11). As the number of recipients receiving a plasmid on the plate is small, the last term in Eq. (D.13) can be disregarded. The equation to be solved is thus

$$\frac{dT_{pk}}{dt} = \gamma_p D_{pk}(t) R_{pk}(t)\,, \tag{D.15}$$

where

$$D_{pk}(t) = D_{k0}\exp(-\lambda t)\,, \tag{D.16}$$
$$R_{pk}(t) = R_{k0}\exp(-\lambda t)\,. \tag{D.17}$$

The solution is

$$T_{pk} = \frac{\gamma_p}{2\lambda} D_{k0} R_{k0}(1-\exp(-2\lambda t))\,, \tag{D.18}$$

which for $t \to \infty$ is

$$T_{pk} = \frac{\gamma_p}{2\lambda} D_{k0} R_{k0}\,. \tag{D.19}$$

The limit for $t \to \infty$ can be considered since the CFU count is made after 12-24 hours, whereafter conjugation on the plate is very unlikely to occur due to death of the donor and recipients. Estimating both γ_p and λ would be an over-parametrization of the model. Therefore γ_p/λ is replaced by the parameter γ'_p. Inserting the result into Eq. (D.11) leads to the observation equation for the so-called *Broth-plate model*

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}_k = \begin{bmatrix} D \\ R \\ T + \frac{\gamma'_p}{2} DR \end{bmatrix}_k + \boldsymbol{e}_k\,,\ \boldsymbol{e}_k \in N(\boldsymbol{0}, \boldsymbol{\Sigma})\,. \tag{D.20}$$

D.3.4 Observation noise

Three different covariance matrices for the observation equation are implemented and compared

Additive:

$$\boldsymbol{\Sigma}_k^a = \begin{bmatrix} s_1^2 & 0 & 0 \\ 0 & s_2^2 & 0 \\ 0 & 0 & s_3^2 \end{bmatrix} \tag{D.21}$$

Proportional:

$$\boldsymbol{\Sigma}_k^p = \begin{bmatrix} s_4^2\sigma_{D,k}^2 & 0 & 0 \\ 0 & s_5^2\sigma_{R,k}^2 & 0 \\ 0 & 0 & s_6^2\sigma_{T,k}^2 \end{bmatrix} \tag{D.22}$$

Additive+Proportional:

$$\boldsymbol{\Sigma}_k^{a,p} = \begin{bmatrix} s_1^2 + s_4^2\sigma_{D,k}^2 & 0 & 0 \\ 0 & s_2^2 + s_5^2\sigma_{R,k}^2 & 0 \\ 0 & 0 & s_3^2 + s_6^2\sigma_{T,k}^2 \end{bmatrix} \tag{D.23}$$

Hence, the observations Y_1, Y_2 and Y_3 are assumed to be uncorrelated in all cases. The variances $\sigma_{R,k}^2$, $\sigma_{R,k}^2$ and $\sigma_{T,k}^2$ are estimates from Eq. (D.3) at time t_k. The most simple noise form is the additive noise, but it is only reasonable, if the variance of the observations is independent of the observations. If this is not the case one alternative is to transform the data to stabilize the variance. In this article instead a proportional noise term is suggested, which include the estimated variance of the observations. Additionally a covariance matrix is considered with both additive and proportional noise.

D.3.5 Statistical methods

The modelling procedure consists of several steps of parameter estimations and goodness of fit statistics. First the Broth model and Broth-plate models are reduced separately, i.e. the likelihood function is optimized for different implementation of the system noise, observation noise and drift term. The best fit for the Broth and Broth-plate models are found applying goodness of fit statistics based on a ML approach. The inference study is made using a likelihood-ratio test and Akaike information criterion, AIC. After reducing the Broth and Broth-plate models they are compared, again using a likelihood-ratio test and AIC. Following the parameter estimation the models are simulated in `Matlab`, by implementing a numerical Euler method as described by Higham (2001).

Parameter estimation

A ML estimation method is used to determine the SDE model parameters. The parameters are found as those maximizing the likelihood function

$$\mathscr{L}(\boldsymbol{\theta};\mathscr{Y}_N) = \left(\prod_{k=1}^{N} p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta})\right) p(\boldsymbol{Y}_0|\boldsymbol{\theta}) \tag{D.24}$$

for the sequence of observations $\mathscr{Y}_N = [\boldsymbol{Y}_N, \boldsymbol{Y}_{N-1}, \ldots, \boldsymbol{Y}_1, \boldsymbol{Y}_0]$. The conditional probability densities p are approximated by gaussian densities motivated by the fact that the SDE (D.4) is driven by a Wiener process having gaussian increments, i.e.

$$p(\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_k^T(\boldsymbol{\Sigma}_{k|k-1}^{yy})^{-1}\boldsymbol{\epsilon}_k\right)}{\sqrt{\det\left(\boldsymbol{\Sigma}_{k|k-1}^{yy}\right)}\left(\sqrt{2\pi}\right)^l} \tag{D.25}$$

where

$$\boldsymbol{\Sigma}^{yy}_{k|k-1} = \mathrm{Var}[\boldsymbol{Y}_k|\mathscr{Y}_{k-1},\boldsymbol{\theta}]\,, \tag{D.26}$$

$$\boldsymbol{\epsilon}_k = \boldsymbol{Y}_k - \hat{\boldsymbol{Y}}_{k|k-1}\,, \text{ and} \tag{D.27}$$

$$\hat{\boldsymbol{Y}}_{k|k-1} = \mathrm{E}[\boldsymbol{Y}|\mathscr{Y}_{k-1},\boldsymbol{\theta}]\,. \tag{D.28}$$

The conditional mean $\hat{\boldsymbol{Y}}_{k|k-1}$ and covariance $\boldsymbol{\Sigma}^{yy}_{k|k-1}$ in the likelihood function (D.24) and (D.25) can be estimated recursively by means of the extended Kalman filter (Kristensen et al., 2004b). In this study we use the software `CTSM`, from which the parameter estimation, correlation of the parameter estimates and the log-likelihood values are obtained. `CTSM` can be downloaded from the webpage: http://www2.imm.dtu.dk/~ctsm/, from where a user manual is also available. `CTSM` is easy to use as it has a built in graphical user interface. Recently an implementation of the method has also been made in R (Klim et al., 2009).

Goodness of fit statistics

Nested models are compared with a likelihood ratio test statistics given by

$$-2\log\Lambda = 2\left(\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0)\right)\,, \tag{D.29}$$

where the test statistics $-2\log\Lambda$ is asymptotically χ^2 distributed with degrees of freedom equal to the difference in dimensions between the two models. $\ell(\boldsymbol{\theta}) = \log(\mathscr{L}(\boldsymbol{\theta};\mathscr{Y}_N))$ and $\ell(\boldsymbol{\theta}_0) = \log(\mathscr{L}(\boldsymbol{\theta}_0;\mathscr{Y}_N))$ are the log-likelihood values of the model and the submodel, respectively. The inference study is also performed with AIC, which is given by

$$AIC = -2\ell(\boldsymbol{\theta}) + 2k\,, \tag{D.30}$$

where k is the number of parameters in the model and $\ell(\boldsymbol{\theta})$ is the log-likelihood value of the model. When comparing models the preferred model is the one with the lowest AIC value.

D.4 Results and discussion

The conjugation experiment continued for 7 h, during which samples were taken from the broth mixture approximately every 20 min and plated on selective agar plates. The CFUs were counted, and the bacterial concentrations in the broth were determined by a generalized linear model approach. The experimental results can be seen in Figure D.2.

D.4.1 Inference study

Log-likelihood values for the Broth model and the Broth-plate model tested with respect to the state noise, the observation noise, and finally the drift term is summarized in Table D.1. When estimating the model parameters only biological plausible parameter intervals are considered. The multiplicative system noise term

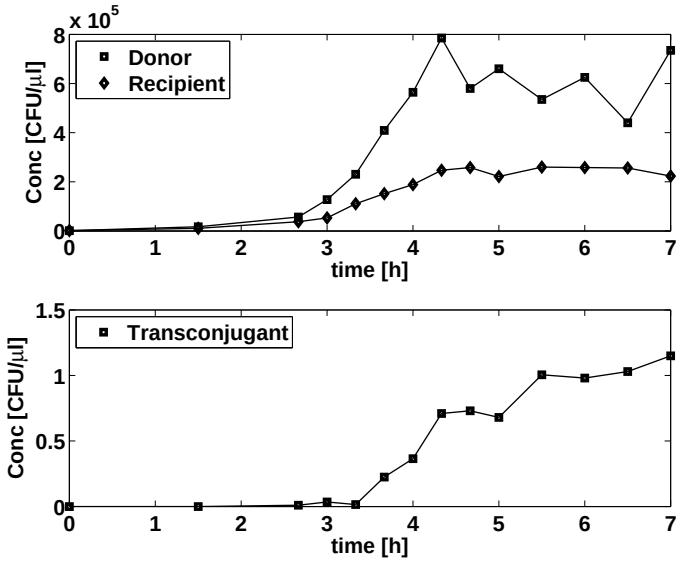


Figure D.2: Data from the conjugation experiment with *E. faecium* growing in broth culture.

Table D.1: Log-likelihood values, p-values and AIC for different versions of the Broth model and Broth-plate model. The nested models listed below each other are tested with the likelihood-ratio test and the p-value for the comparison is listed in the line of the smallest of the two models tested. The proportional and additive observation noise structures are both tested against the proportional+additive noise term.

	Broth model			Broth-plate model		
	$\ell(\hat{\boldsymbol{\theta}})$	p-value	AIC	$\ell(\hat{\boldsymbol{\theta}})$	p-value	AIC
System noise:						
Multiplicative	-299.04	-	638.08	-298.05	-	634.09
Additive	-299.19	-	634.38	-298.05	-	634.09
Additive ($\sigma_D = \sigma_R = \sigma_T$)	-299.19	0.9999	630.38	-298.05	0.9999	630.09
Observation noise:						
Proportional+additive	-299.19	-	630.38	-298.05	-	630.09
Proportional	-300.52	0.4467	627.04	-299.47	0.4162	626.94
Additive	-312.67	0.0006	651.34	-311.43	0.0007	650.86
Drift term:						
$\alpha = 0$	-300.52	0.9969	625.04	-299.47	0.9695	624.94
$\kappa_T = \kappa_R$	-300.59	0.7014	623.19	-299.53	0.7232	623.06

(D.9) is tested against additive system noise (D.10), and it is found that there is no significant difference. We therefore continue the study for both the Broth and Broth-plate model with additive system noise. It would be interesting to investigate the system noise structure with a ML estimation based on the particle filter (Ionides et al., 2006), for which the multiplicative noise can be implemented directly dependent on the state variable. We will leave this for a future study. The

additive system noise is well modelled with $\sigma = \sigma_D = \sigma_R = \sigma_T$ for both the Broth and Broth-plate models.
Both the proportional observation noise and the additive observation noise can be tested against the proportional+additive observation noise using a likelihood-ratio test. The model with additive observation noise is seen to perform significantly worse than the proportional+additive noise model (p-value = 0.0006 and p-value = 0.0007), whereas data is well modelled with proportional observation noise (p-value = 0.4467 and p-value = 0.4162). All the observation noise covariance matrices can be compared using AIC, from which it is also found that data is best modelled with proportional observation noise. This result is as expected, as the variance of the observations increases for higher CFU counts, which is best captured by proportional noise.
With this noise structure the two hypothesis $\alpha = 0$ and $\kappa_T = \kappa_R$ are tested. We fail to reject both hypothesis as the p-values are high (between 0.7014 and 0.9969).
The best description of data for each of the models is thus a model with additive system noise where the standard deviation is the same for D, R and T. The observation noise should be modelled as proportional noise and growth for the transconjugants and recipients can be modelled with the same set of parameters.
The reduced Broth and Broth-plate models can be compared with the likelihood-ratio test, as they are nested models with the Broth-plate model containing one additional parameter (γ'_p) compared to the Broth model. The test statistics for this test is 2.12, which gives a p-value of 0.1451. This is not significant on ordinary level and thus the smaller Broth model should be chosen. However, this conclusion is based on very few datapoints, and it might change if more data were available.

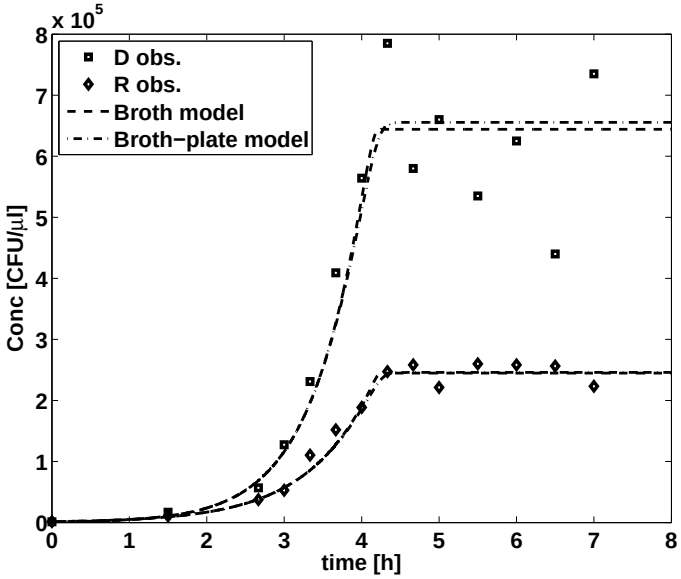
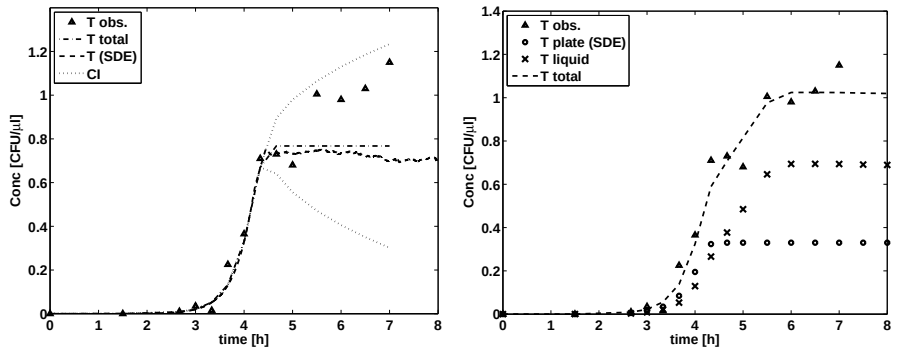


Figure D.3: The donor and recipient concentration over time as simulated by the Broth and Broth-plate models plotted together with data values.

Table D.2: The result of the ML parameter estimation for the Broth model and the Broth-plate model.

Parameter	**Broth model** (SD)		**Broth-plate model** (SD)	
υ_D [h^{-1}]	1.678	(0.113)	1.787	(0.092)
$\upsilon_R = \upsilon_T$ [h^{-1}]	1.216	(0.028)	1.261	(0.035)
κ_D	0.052	(0.047)	0.110	(0.047)
$\kappa_R = \kappa_T$	0.0034	(0.0032)	0.0364	(0.0199)
η	$1.128{\cdot}10^{-6}$	$(0.079 \cdot 10^{-6})$	$1.112{\cdot}10^{-6}$	$(0.087{\cdot}10^{-6})$
κ_c	$4.201{\cdot}10^{-16}$	$(3.966 \cdot 10^{-16})$	$5.583{\cdot}10^{-15}$	$(8.811{\cdot}10^{-15})$
$\gamma_{\max}$ [(CFU/μl)$^{-1}$h^{-1}]	$4.913{\cdot}10^{-12}$	$(0.636{\cdot}10^{-12})$	$2.008{\cdot}10^{-12}$	$(0.699{\cdot}10^{-12})$
γ'_p [(CFU//μl)$^{-1}$]	0	-	$4.096{\cdot}10^{-12}$	$(1.630{\cdot}10^{-12})$
σ	$4.010 \cdot 10^{-10}$	$(2.828 \cdot 10^{-10})$	$1.128 \cdot 10^{-11}$	$(2.488 \cdot 10^{-11})$
σ_4	13.815	(5.349)	14.862	(5.575)
σ_5	3.507	(1.212)	3.572	(1.304)
σ_6	3.150	(1.422)	3.184	(1.280)



(a) Broth model (b) Broth-plate model

Figure D.4: The transconjugant concentration plotted together with simulations of the Broth and Broth-plate models. For the Broth model the mean of the simulated transconjugant concentration and its confidence interval, as estimated by `CTSM`, is shown. For each model one example of the SDE simulation made in `Matlab` is plotted.

D.4.2 Parameters

The result of the ML parameter estimation for the reduced Broth and Broth-plate models is shown in Table D.2, and the results of simulations with these parameters are plotted in Figure D.3 and D.4.

The conjugation rate is estimated to $2.008 \cdot 10^{-12}$ (CFU/μl)$^{-1}$h^{-1} for the Broth-plate model and $4.913 \cdot 10^{-12}$ (CFU/μl)$^{-1}$h^{-1} for the Broth model. Thus, the choice of model influence the estimate of the conjugation rate.

The value of κ_c ($10^{-16} - 10^{-15}$) is low compared to the κ value for the growth of the bacteria ($10^{-3} - 10^{-2}$). This means that the mass action model γDR in this con-

jugation experiment gives a good description of conjugation until the substrate is depleted. As a consequence of the low κ_c value the model predicts conjugation to continue after the growth of the bacteria has reached stationary phase. By comparing the simulations of the donor and recipient concentrations Figure D.3 with simulation of the transconjugant concentration Figure D.4 we see that new transconjugants appear until approximately one hour after initiation of the stationary growth phase.
When calculating the bacterial concentration from the CFU count it was assumed that the data were Poisson distributed, and thus that the variance equals the mean. If this was indeed the case each of the parameters σ_4^2, σ_5^2, and σ_6^2 should equal one. In this case σ_4^2 is around 14 and σ_5^2 and σ_6^2 are around 3. This means that the data are over-dispersed, and this over-dispersion is accounted for by σ_4^2, σ_5^2, and σ_6^2. These parameters are in Poisson regression also refer red to as dispersion parameters. Several authors have found asymmetric likelihood profiles (Dennis et al., 2006; Ionides et al., 2006; Ponciano et al., 2007; King et al., 2008) in state-space models for dynamical biological systems (e.g. bird growth, plasmid persistence and cholera pandemic). In order to check the likelihood structure for our conjugation model, the profile-likelihood is calculated for ν_D and $\gamma_{\max}$. The profile likelihoods seen in Figure D.5 are calculated by optimizing the likelihood function for fixed values of the parameter of interest. The 95% confidence interval (CI) plotted is the region of parameter values for which the profile log-likelihood value is larger than $\ell(\boldsymbol{\theta})_{\max} - c/2$, where $\ell(\boldsymbol{\theta})_{\max}$ is the maximum log-likelihood value and c is defined by $\text{Prob}[\chi^2(1) < c] = 0.95$. The profile likelihoods are indeed asymmetric, which also in 3 out of 4 cases leads to asymmetric CIs. This is as expected due to the small number of observations. Furthermore it is seen that the profile likelihood CI is generally wider than the CI calculated by `CTSM`. However, this difference is not very large and we therefore believe that it has no or only a limited influence on the inference study performed. It should be noted that the observation noise seems to increase when the profile likelihood value is decreasing. This indicates that the observation noise (and not as expected the system noise) explains the difference between observations and the model. This is due to the few observations, which makes it difficult to adequately separate observation and system noise as discussed by Dennis et al. (2006).
In addition to the samples directly plated on the transconjugant plates, also 10 times concentrated samples were plated for the first six time points, where transconjugant concentration in the flask was low. However, an expected 100 times increase in conjugation on the plate compared to the non-concentrated sample was not observed. The reason for this is not clear. The sample is centrifuged in order to make a concentrated sample, this might change the ability for the donor and recipient to conjugate on the plate which could be one explanation. Further experiments should be performed to support the finding of conjugation under antibiotic pressure on the selective agar plate.

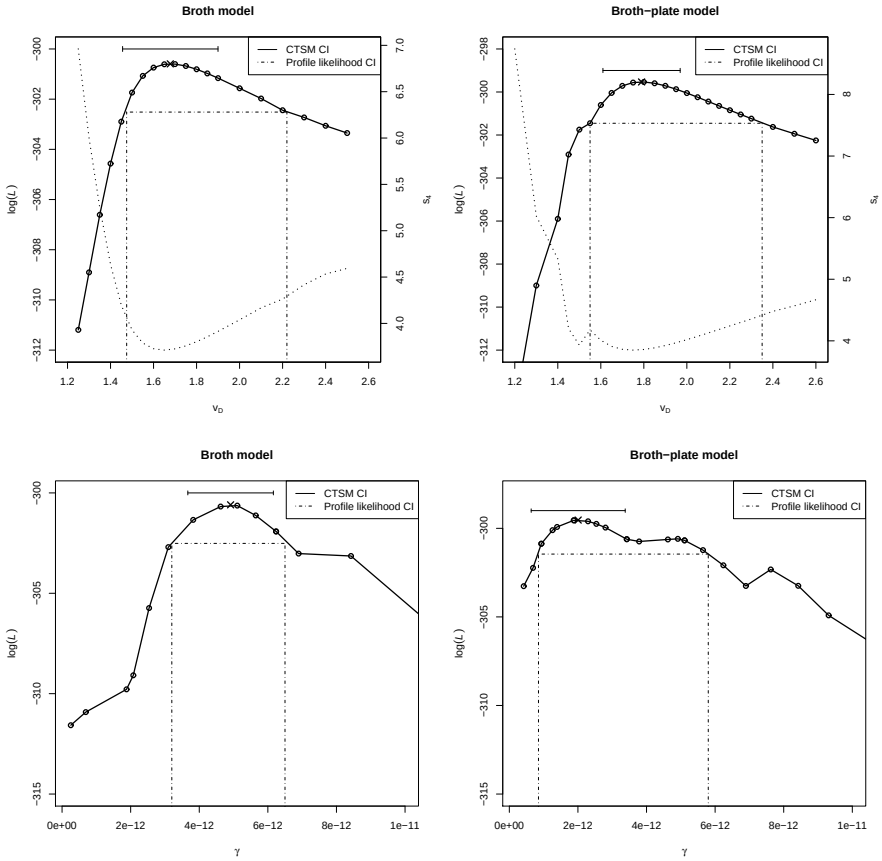


Figure D.5: Profile likelihood (full line and circles) and approximate 95% confidence interval for v_D and $\gamma_{\max}$. The cross marks the estimated parameter value. The dotted lines in the top plots give the standard deviation s_4.

D.5 Conclusions

The proposed SDE based state-space model is shown to successfully model conjugation in a broth exhaustible media. The suggested substrate dependent expression for the conjugation rate satisfactory model the conjugation process, which stops as the substrate is depleted. The ML based framework for estimating model parameters combined with likelihood-ratio tests and AIC for inference studies provides strong tools for model improvements. It is shown that the stochasticity of the observations is best modelled as proportional noise.
The methodological problem of conjugation occurring on the transconjugant plates motivates the development of the Broth-plate model, which includes plate conjugation in the observation equation, whereby the plate conjugation can be separated from conjugation in the broth media. However, for the given data the Broth-plate model does not perform significantly better than the Broth-model.

Acknowledgements

This study was supported by the Danish Research Council for Technology and Production Sciences through the Grant 274-05-0117 titled "Evolution and adaptation of antimicrobial resistance in bacterial populations".

D.6 Bibliography

Andrup, L., Andersen, K., 1999. A comparison of the kinetics of plasmid transfer in the conjugation systems encoded by the f plasmid from *Escherichia coli* and plasmid pcf10 from *Enterococcus faecalis*. Microbiology 145 (8), 2001–2009.

Andrup, L., Smidt, L., Andersen, K., Boe, L., 1998. Kinetics of conjugative transfer: A study of the plasmid pxo16 from *Bacillus thuringiensis* subsp. *israelensis*. Plasmid 40 (1), 30–43.

Clewlow, L. J., Cresswell, N., Wellington, E. M. H., 1990. Mathematical model of plasmid transfer between strains of streptomycetes in soil microcosms. Appl. Environ. Microbiol. 56 (10), 3139–3145.

De Gelder, L., Ponciano, J. M., Abdo, Z., Joyce, P., Forney, L. J., Top, E. M., 2004. Combining mathematical models and statistical methods to understand and predict the dynamics of antibiotic-sensitive mutants in a population of resistant bacteria during experimental evolution. Genetics 168 (3), 1131–1144.

De Gelder, L., Ponciano, J. M., Joyce, P., Top, E. M., 2007. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiology (Reading, Engl.) 153 (Part 2), 452–463.

Dennis, B., Ponciano, J. M., Lele, S. R., 2006. Estimating density dependence, process noise, and observation error. Ecological Monographs 76 (3), 323.

Freter, R., Freter, R. R., Brickner, H., 1983. Experimental and mathematical models of *Escherichia coli* plasmid transfer in vitro and in vivo. Infect. Immun. 39 (1), 60–84.

Hasman, H., Aarestrup, F. M., 2002. Mechanisms of resistance - *tcrB*, a gene conferring transferable copper resistance in *Enterococcus faecium*: Occurrence, transferability, and linkage to macrolide and glycopeptide resistance. Antimicrob. Agents Chemother. 46 (5), 1410.

Higham, D. J., 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM review 43 (3), 525–546.

Ionides, E. L., Bretó, C., King, A. A., 2006. Inference for nonlinear dynamical systems. Proc. Natl. Acad. Sci. U. S. A. 103 (49), 18438–18443.

Joyce, P., Abdo, Z., Ponciano, J. M., Gelder, L. D., Forney, L. J., Top, E. M., 2005. Modeling the impact of periodic bottlenecks, unidirectional mutation, and observational error in experimental evolution. J. Math. Biol. 50 (6), 645–662.

King, A. A., Ionides, E. L., Pascual, M., Bouma, M. J., 2008. Inapparent infections and cholera dynamics. Nature 454 (7206), 877–880.

Klim, S., Mortensen, S. B., Kristensen, N. R., Overgaard, R. V., Madsen, H., 2009. Population stochastic modelling (psm) - an r package for mixed-effects models based on stochastic differential equations, computer methods and programs in biomedicine, in press. Comput. Methods Programs Biomed. 94 (3), 279–289.

Knudsen, G. R., Walter, M. V., Porteous, L. A., Prince, V. J., Armstrong, J. L., Seidler, R. J., 1988. Predictive model of conjugative plasmid transfer in the rhizosphere and phyllosphere. Appl. Environ. Microbiol. 54 (2), 343–347.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004a. A method for systematic improvement of stochastic grey-box models. Comput. Chem. Eng. 28, 1431–1449.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004b. Parameter estimation in stochastic gray-box models. Automatica 40 (40), 225–237.

Levin, B., F.M., S., Rice, V., 1979. The kinetics of conjugative plasmid transmission: fit of a simple mass action model. Plasmid 2 (2), 247–260.

Levin, B. R., Stewart, F. M., 1980. The population biology of bacterial plasmids: A priori conditions for the existence of mobilizable nonconjugative factors. Genetics 94 (2), 425–443.

MacDonald, Jacqueline, A., Smets, Barth, F., Rittmann, Bruce, E., 1992. Effects of energy availability on the conjugative-transfer kinetics of plasmid rp4. Water Res. 26 (4), 461–468.

Monod, J., 1949. The growth of bacterial cultures. Annu. Rev. Microbiol. 3, 371–394.

Ponciano, J. M., Gelder, L. D., Top, E. M., Joyce, P., 2007. The population biology of bacterial plasmids: A hidden markov model approach. Genetics 176 (2), 957.

Slater, F. R., Bailey, M. J., Tett, A. J., Turner, S. L., 2008. Progress towards understanding the fate of plasmids in bacterial communities. FEMS Microbiol. Ecol. 66 (1), 3–13.

Sudarshana, P., Knudsen, G. R., 2006. Quantification and modeling of plasmid mobilization on seeds and roots. Curr. Microbiol. 52 (6), 455–459.

Tier, C., Floyd, H., 1981. Persistence in denisty dependent stochastic populations. Math. Biosci. 53, 89–117.

Top, E., Vanrolleghem, P., Mergeay, M., Verstraete, W., 1992. Determination of the mechanism of retrotransfer by mechanistic mathematical modeling. J. Bacteriol. 174 (18), 5953–5960.

Willms, A., Roughan, P., Heinemann, J., 2006. Static recipient cells as reservoirs of antibiotic resistance during antibiotic therapy. Theor. Popul. Biol. 70 (4), 436–451.

D.A Generalized linear model estimation

For any given observation time t_k, the concentration Y_k in the flask of a given bacterial population can be found from the following `Matlab` code:

```
% Create a vector N_k with the CFU counts at time t_k, e.g.
   N_k = [94, 152, 8, 5, 1, 3];

% Create a vector n with the dilution for each of the
% observations N_k at time t_k, e.g
   n = [0.1, 0.1, 0.01, 0.01, 0.001, 0.001];

   x0 = ones(length(N_k),1);
   X = log(n);

% Estimating the parameter, beta, the deviance of the
% estimation, dev, and the statistics for the test
% (including standard deviation).

   [beta,dev,stat] = glmfit(x0,N_k,'poisson','link','log',...
                            'offset',X,'constant','off');

% The concentration at time t_k is
   Y_k = exp(beta);

% The standard deviation at time t_k is
   sigma_Yk = Y_k*stat.se;
```

PAPER

E

Mathematical model for competitive growth of *P. aeruginosa* and mutator strains in sub-MIC concentration of ciprofloxacin ‡

Abstract

Pseudomonas aeruginosa mutators characterized by a high mutation rate are found with high frequencies in the lungs of cystic fibrosis patients, and they are believed to play an important role for the evolution of resistance. In this study a mathematical model is suggested to describe the dynamics of competitive growth between *P. aeruginosa* wild-type and the *mutM*, *mutY*, *mutT* and *mutY-mutM* mutators. When growing in sub-MIC concentrations of ciprofloxacin (0.1 μg/ml, MIC = 0.125-0.19), the mutator population is found to gradually take up a larger part of the total population. This is caused by the mutators ability to adapt to the stressed environment faster, and thereby obtain a higher fitness. The model suggests that the changes obtained by the mutators to achieve a higher fitness are not mutations, as the rate of adapting to the environment is much higher than the estimated mutation rate. The equilibrium concentration of mutators is found to be dependent on the strength of the mutator as compared to the wild-type bacteria.

Pseudomonas aeruginosa causes very critical and complicated infections, for which therapy is strongly dependent on successful antibiotic treatment. In cystic fibrosis (CF) patients, chronic lung infection caused mainly by *P. aeruginosa*

‡Manuscript under preparation: K. R. Philipsen, L. E. Christiansen, L. F. Mandsberg, O. Ciofu, H. Madsen. Mathematical model for competitive growth of *P. aeruginosa* and mutator strains in sub-MIC concentration of ciprofloxacin

is the major cause of death (Smith and Travis, 1996; Gibson et al., 2003; Lyczak et al., 2002). Intensive antibiotic treatment has improved the survival and clinical condition of CF patients, however, the development of resistance to antibiotics makes the treatment difficult. Mutators (also called hypermutators) characterized by high mutation rates have been found with high frequencies in CF patients (Oliver et al., 2000; Ciofu et al., 2005). Several studies have shown that the frequency of mutators increase due to hitchhiking with adaptive mutations (Mena et al., 2008; Giraud et al., 2001; Mao and Lane, 1997; Miller and Suthar, 1999). These findings have been supported by several mathematical models (Tenaillon and Toupance, 1999; Travis and Travis, 2002; Tanaka et al., 2003; Travis and Travis, 2004). Thus, mutators are believed to play an important role for the evolution of resistance (Oliver et al., 2000; Maciá et al., 2006). Recently Ferroni et al. (2009) have shown that bacteria strains from CF patients with a higher mutation rate increase the rate of acquisition of new antibiotic resistance. In addition, Mena et al. (2008) have found that mutators also play an important role in the evolution and adaptation of the bacteria population during chronic respiratory infections.
Different studies have been performed to investigate the fitness of mutator bacteria relative to the wild-type (non-mutator) strains. In an experiment with mice, Maciá et al. (2006) found a higher concentration of the *mutS* mutator bacteria than the wild-type bacteria after treatment with ciprofloxacin. Ciprofloxacin is a commonly used antibiotic for the initial treatment of *P. aeruginosa* infections in CF patients. Another competition experiment between wild-type *P. aeruginosa* and the laboratory *mutS* mutator has been performed by Montanari et al. (2007). After 24 hours of competing growth they obtained a significantly higher wild-type concentration than mutator concentration. However, this result was influenced by a lower growth rate of the mutator bacteria, giving it a disadvantage in the competition with the wild-type.
In recent studies, laboratory *P. aeruginosa mutT*, *mutY*, *mutM* and *mutY-mutM* mutants have been constructed and characterized (Mandsberg et al., 2009, 2010). As opposed to the strains used by Montanari et al. (2007) the growth rates of these strains have been found to be equal when growing in Luria-Bertani (LB) media (Philipsen et al., 2008). A competition experiment between PAO1 and the *mutY-mutM* mutant in LB and LB with sub-MIC concentration of ciprofloxacin has recently been described in Mandsberg et al. (2010). The study showed that the mutator strains take over the population after three days of competing growth in sub-MIC concentrations of ciprofloxacin.
Several earlier simulation studies have been performed to theoretically investigate the frequency of mutator bacteria in a competing environment (**?**Tenaillon and Toupance, 1999; Tanaka et al., 2003; Travis and Travis, 2002, 2004; Pal et al., 2007). These studies all uses discrete time simulation and are either deterministic (Travis and Travis, 2002, 2004) or stochastic with Poisson sampling or Binomial Sampling (**?**Tenaillon and Toupance, 1999; Tanaka et al., 2003; Pal et al., 2007). To our knowledge we are the first to apply a continuous time stochastic model in the form of stochastic differential equations to simulate competitive growth between

a mutator and wild-type strains, and to estimate the model parameters based on experimental data. The mathematical model will be developed from the experimental results and subsequently used to discuss the dynamics of the system.

E.1 Materials and methods

E.1.1 Growth and competition experiment

The strains included in this study have been described in previous articles (Mandsberg et al., 2009, 2010). The reference strain used is PAO1 from Stover et al. (2000). Competition experiments were carried out in a bioscreen (Labsystem C, Bie og Berntsen) 2x10 wells with LB-media or with LB-media with 0.1 μg/ml ciprofloxacin (MIC = 0.125-0.19). We attempted to start with a ratio of 1:1 for PAO1 and each of the four mutator strains *mutT*, *mutY*, *mutM*, *mutY-mutM* in each well. Growth of PAO1, and each of the four mutator strains *mutT*, *mutY*, *mutM mutY-mutM* were measured in 2x1 wells with Luria-Bertani (LB) media and with LB-media with 0.1 μg/ml ciprofloxacin simultaneously with the competition experiment. The inoculum were cultures diluted to 10^{-2} then adjusted to approx OD600=0.035 and then further diluted 1000 times. 140 μl of each diluted culture was mixed in microtitterwells and the growth was carried out at 37°C, continuously shaking and taking optical density (OD) measurements every 30 min for 24 hours. The start colony forming units (CFU) was determined on LB plates. For five consecutive days (start day 0, end day 4) the up grown culture was diluted 1000 times and transferred to a new microtitterplate for exponential growth throughout the experiment. Each day the up grown cultures from the competition experiment were serial diluted and plated on LB agar and on LB agar supplemented with 30 μg/ml gentamycin. The mutator strains were not inhibited by this concentration of gentamycin and the actual CFU of PAO1 and the mutators were calculated; the mutators on gentamycin plates and PAO1 and the mutators on LB plates. A simple subtraction then gave the CFU for PAO1. The ratio of PAO1:*mutM*, PAO1:*mutY*, PAO1:*mutT* and PAO1:*mutY-mutM* were followed over the five days.

E.1.2 Conversion of OD measurements to cell concentration

The mean media contribution was subtracted from the OD measurements to obtain only the bacterial growth. Subsequently, the OD measurements were converted to cell concentration by applying the exponential calibration curve

$$Conc = -\frac{1}{b}\log(1 - \frac{OD}{a}), \tag{E.1}$$

introduced by Philipsen et al. (2010b) and using the constants a and b as estimated in that study.

Table E.1: Mutation rates to ciprofloxacin estimated by a fluctuation experiment and the related mutator strength, i.e. the factor by which the mutation rate for the mutator strain is increased in comparison to the wild-type bacteria.

Bacteria	Mutation rate [per cell per generation]	mutator strength
PAO1	$3.153 \cdot 10{-}9$	
mutM mutant	$4.781 \cdot 10{-}9$	1.5
mutT mutant	$4.281 \cdot 10{-}8$	13.6
mutY mutant	$9.845 \cdot 10{-}8$	31.2
mutM-mutY mutant	$2.498 \cdot 10{-}7$	79.2

E.1.3 Modelling framework

A Stochastic Differential Equation (SDE) framework (Kristensen et al., 2004; Philipsen et al., 2010a) was used to model the bacteria dynamics. The use of SDEs enables several strong tools for parameter estimation and model development (Kristensen et al., 2004; Philipsen et al., 2010c). The software `CTSM` (www.imm.dtu.dk/~ctsm) was used to estimate the model parameters and obtain information about the standard deviation of the estimated parameters.

E.2 Results

E.2.1 Experimental results

Prior to the experiment we ensured that PAO1 and the mutators have equal growth rates in the initial exponential growth phase in LB (p=0.607) (Philipsen et al., 2008) and LB with ciprofloxacin (p=0.050), using a method described in Philipsen et al. (2008). Furthermore the mutation rates were calculated based on a fluctuation experiment as described in Mandsberg et al. (2009). The mutation rate, μ, for each bacteria strain is given in Table E.1. As expected, the PAO1 *mutM-mutY* mutant is the strongest mutator with a mutator strength of 79.2, e.g. a mutation rate which is 79.2 times higher than the rate for PAO1. The mutation rate for PAO1 is 10 times lower than the rate estimated by Maciá et al. (2006), but in accordance with previously estimated mutation rates to rifampicin (Mandsberg et al., 2009).

From the experiment with growth of single bacteria strains (Figure E.1) it is seen that the sub-MIC concentration of ciprofloxacin results in a longer lag phase and a longer diauxic lag as compared to the growth in LB for all the bacteria strains. Figure E.2 shows the results of the competition experiment for growth of PAO1 and each of the four mutator strains. As expected, the pattern of the growth is very similar to that seen for the growth of each bacterial strain individually.

The bacterial concentration in the end of each day, as estimated from OD measurements (dotted lines) and the plate counts (triangles), is very similar for growth in LB. This shows that the exponential calibration curve makes a good translation

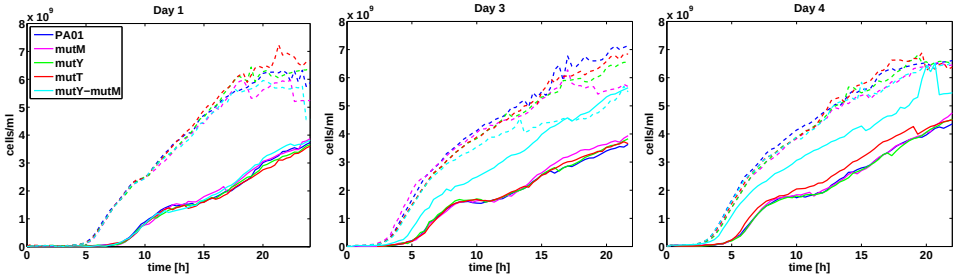


Figure E.1: The concentration of each of the five bacteria strains when growing in LB (dotted line), and LB with ciprofloxacin (full line). On day two the lamp in the bioscreen did not work, and there are therefore no measurements for this day.

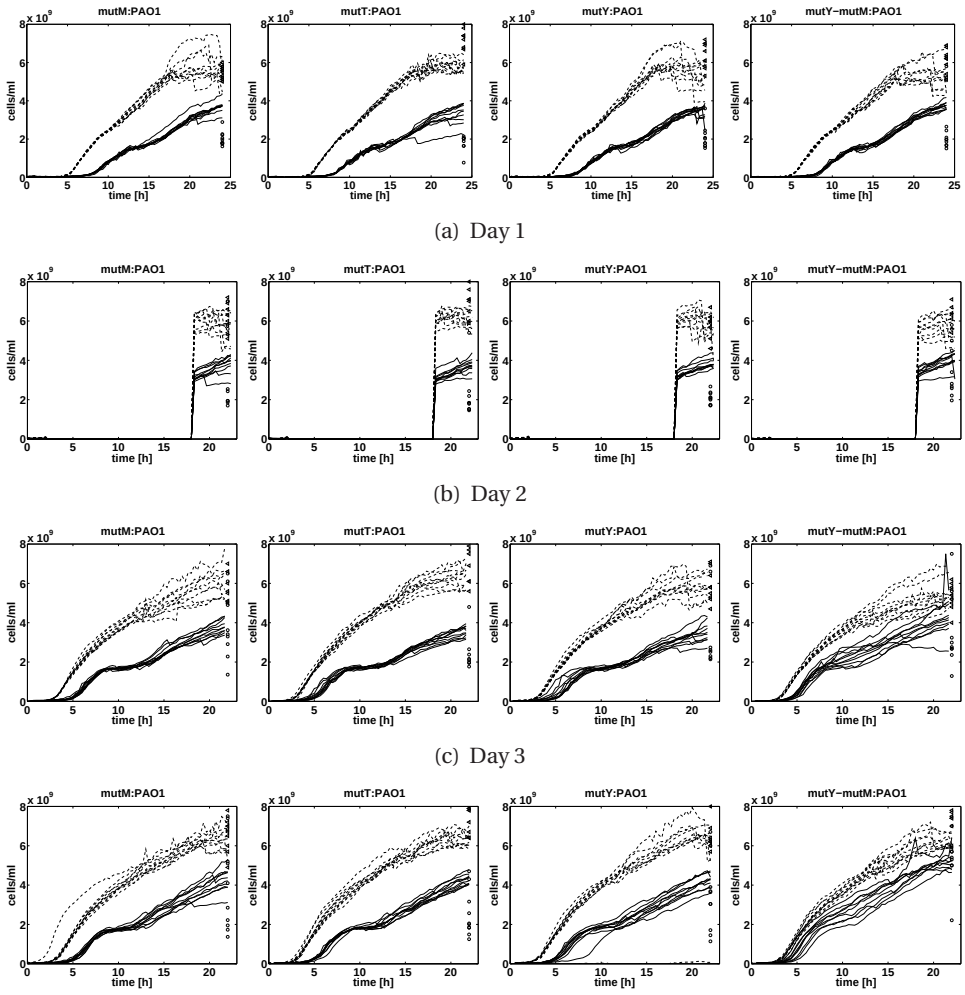
from the OD measurements to bacterial concentrations. For growth in LB with ciprofloxacin, however, the plate counts are consistently smaller than the corresponding OD measurements. The plate counts and OD measurements are only the same for *mutY-mutM*:PAO1 on day 4. We believe that the difference observed for all the other days is due to filament structure in the bacterial populations, which can lead to a lower CFU count (Blázquez et al., 2006). The mathematical model will therefore be built on the result of the OD measurements.

From plate counts, the percentage of mutator bacteria in the population is calculated for each day (Figure E.3). For the *mutY-mutM*:PAO1 competition experiment, the mutator is seen to gradually overtake the population when growing in LB with ciprofloxacin, whereas the ratio between mutator and PAO1 stays constant for all days for growth in LB. This result was also reported by Mandsberg et al. (2010). For the *mutM*:PAO1 competition experiment there are indications that the mutator bacteria is present in higher concentrations than PAO1 on day 3 and 4. There is no clear increase in the mutator percentage when growing in LB with ciprofloxacin for the other experiments.

E.2.2 Mathematical model

A mathematical model is developed based on the experimental results to describe the growth on LB and LB with ciprofloxacin. The growth of wild-type and mutator PAO1 on LB media and LB with ciprofloxacin is diauxic as seen from Figure E.1. This dynamic can be described using the so-called optimal (or cybernetic) model (Doshi et al., 1997; Kompala et al., 1984; Bajpai-Dikshit et al., 2003). We use a growth rate proportional to the substrate content as proposed by Philipsen et al. (2010c). The bacteria is assumed to first grow on substrate S_1 which is facilitated by enzyme E_1 and subsequently on substrate S_2 induced by enzyme E_2. The transition between the substrates is regulated by the fraction α,

$$\alpha = \frac{(1+k)S_1^2}{k+S_1^2}, \tag{E.2}$$



(a) Day 1

(b) Day 2

(c) Day 3

(d) Day 4

Figure E.2: The total concentration of PAO1 and each of the four mutator strains mutM, muT, mutY and mutYM when growing in LB (dotted line) and LB with ciprofloxacin (full line). The triangles and circles mark the total bacterial concentration for growth in LB and LB with ciprofloxacin, respectively, as measured by plate counts in the end of each day.

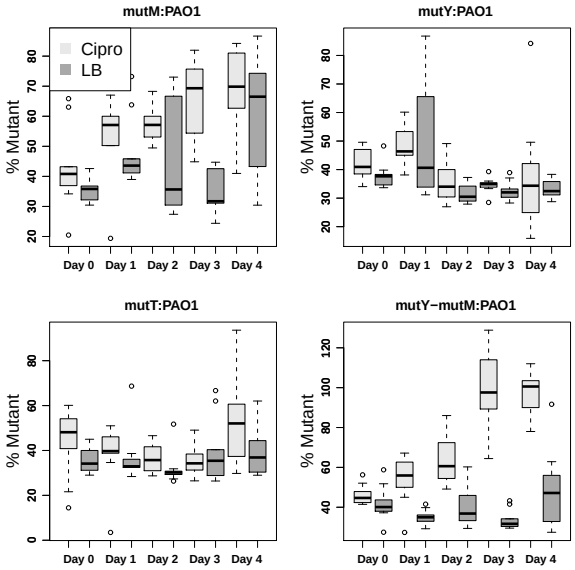


Figure E.3: The percentage of the bacterial population which consists of mutators for the initial population (Day 0) and each of the following four days of the experiment.

where k determines for which level of S_1 the activation of enzyme E_2 begins and the term $(1+k)$ ensures that α changes from 1 (for which only substrate S_1 is used) to 0 (for which substrate S_2 is fully utilized). S_1 and S_2 are implemented as normalized states with initial values of 1. The diauxic lag is longer for growth in LB with ciprofloxacin than for growth in LB, which means that the activation of the second enzyme takes longer when ciprofloxacin is present. This effect can be captured in the model by a difference in the value of k under the two growth conditions.
There is a longer lag phase on day one for growth in both LB and LB with ciprofloxacin as compared to the subsequent days. This indicates that the bacteria need to adapt to the new environment before growth can begin. This effect can be modelled by including an additional adapted state, A, to the model and assuming that a bacteria first can start growing when it becomes adapted. For growth in LB the transfer to the adapted population is only modelled on day 1. On the subsequent days the bacteria can be assumed to be adapted from the beginning of the growth. Using SDEs the model for growth in LB is described by the following set of equations

$$\begin{aligned}
dB &= -\gamma B\,dt + \sigma_B\,d\omega \\
dA &= [\gamma B + (\upsilon_1 S_1 E_1 + (1-\alpha)\,\upsilon_2 S_2 E_2)\,A]\,dt + u_A \sigma_A\,d\omega \\
dE_1 &= [(\upsilon_1 + \beta) S_1 - \upsilon_1 S_1 E_1 E_1 - \beta E_1]\,dt + \sigma_E\,d\omega \\
dE_2 &= [(1-\alpha)[(\upsilon_2 + \beta) S_2 - \upsilon_2 S_2 E_2 E_2] - \beta E_2]\,dt + \sigma_E\,d\omega \\
dS_1 &= -\eta_1 \upsilon_1 S_1 E_1 A\,dt + \sigma_S\,d\omega \\
dS_2 &= -(1-\alpha)\eta_2 \upsilon_2 S_2 E_2 A\,dt + \sigma_S\,d\omega
\end{aligned} \tag{E.3}$$

The first term in each equation is called the drift term, and the second term is called the diffusion term. The drift term describes the known dynamics of the system. For example the bacteria, B, can become adapted with the rate γ, and the adapted population grows on substrate S_1 with the rate $\upsilon_1 S_1 E_1$ and on substrate S_2 with the rate $(1-\alpha)\upsilon_2 S_2 E_2$. The diffusion term accounts for noise in the system and for those biological mechanisms not included in the model. The parameters in the drift term are the growth rates, υ_1 and υ_2, the yield factors, $1/\eta_1$ and $1/\eta_2$, the rate γ by which the bacteria adapt to the new environment and the rate β by which the enzyme is removed from the population. In the diffusion term, $d\omega$ is the increments of a standard Wiener process and σ_E, σ_S and σ_B are the standard deviation for the increments of the Wiener process. For the adapted state, A, a multiplicative system noise is implemented by including an input u_A which is the observed bacterial concentration. The growth parameters are the same for all bacteria strains and they are estimated based on the growth data for day 1. The yield factors seem to change over time, and new values for η_1 and η_2 are therefore estimated from the growth data from day 3, while keeping the remaining parameters fixed.

When the bacteria grow in ciprofloxacin we assume that they can become more fit to the environment by adaptation. We will call this new population resistant, R, even though it will be clear later that the rate of "mutation" to the "resistant" population is much higher than the mutation rate listed in Figure E.1. Indexing the parameters for the resistant population with an $*$ the growth model for growth in LB with ciprofloxacin becomes

$$\begin{aligned}
dB &= -\gamma B\,dt + \sigma_B\,d\omega \\
dA &= [\gamma B + (1-\mu)(\upsilon_1 S_1 E_1 + (1-\alpha)\upsilon_2 S_2 E_2)A]\,dt + u_A\sigma_A\,d\omega \\
dR &= [\mu(\upsilon_1 S_1 E_1 + (1-\alpha)\upsilon_2 S_2 E_2)A + \\
&\quad (\upsilon_1^* S_1 E_1^* + (1-\alpha^*)\upsilon_2^* S_2 E_2^*)R]\,dt + u_R\sigma_R\,d\omega \\
dE_1 &= [(\upsilon_1+\beta)S_1 - \upsilon_1 S_1 E_1 E_1 - \beta E_1]\,dt + \sigma_E\,d\omega \\
dE_2 &= [(1-\alpha)[(\upsilon_2+\beta)S_2 - \upsilon_2 S_2 E_2 E_2] - \beta E_2]\,dt + \sigma_E\,d\omega \\
dE_1^* &= [(\upsilon_1^*+\beta^*)S_1 - \upsilon_1^* S_1 E_1^* E_1^* - \beta^* E_1^*]\,dt + \sigma_E\,d\omega \\
dE_2^* &= [(1-\alpha^*)[(\upsilon_2^*+\beta^*)S_2 - \upsilon_2^* S_2 E_2^* E_2^*] - \beta^* E_2^*]\,dt + \sigma_E\,d\omega \\
dS_1 &= [-\eta_1\upsilon_1 S_1 E_1 A - \eta_1^*\upsilon_1^* S_1 E_1^* R]\,dt + \sigma_S\,d\omega \\
dS_2 &= [-(1-\alpha)\eta_2\upsilon_2 S_2 E_2 A - (1-\alpha^*)\eta_2^*\upsilon_2^* S_2 E_2^* R]\,dt + \sigma_S\,d\omega
\end{aligned} \tag{E.4}$$

where

$$\alpha^* = \frac{(1+k^*)S_1^2}{k^* + S_1^2}\,. \tag{E.5}$$

The parameter estimation for the growth in LB with ciprofloxacin is carried out in three steps. 1) First the model without resistant bacteria is considered (similar to the model for growth in LB), and the parameters are estimated from the data for

day 1 (Figure E.1 left). This is possible as the number of resistant bacteria arising during the first day can be assumed to be very small. The rate of adaptation for the non-resistant population from day 2 is different than for the first day. This change in rate is estimated from the growth data of PAO1 for day 3 (as the number of resistant bacteria for this population can be assumed to be very low) while keeping all other rates fixed. 2) Second, the growth parameters for the resistant population are estimated. From the competition experiment we know that the *mutY-mutM* mutator has taken over the population at day 4 (Figure E.3). It can therefore be assumed that the *mutY-mutM* population at day 4 consists only of resistant bacteria. The growth parameters for the resistant bacteria can thus be estimated from the growth curve of *mutY-mutM* on day 4 (Figure E.1 right) by using the model for growth in LB and replacing the adapted population with a resistant population. 3) Finally the mutation rate should be estimated. The mutation rate for *mutY-mutM* is fitted by eye, such that the simulation of the growth model in LB with ciprofloxacin follows the result from the growth experiment. The mutation rate of PAO1 and each of the three other mutator strains are adjusted accordingly to the difference between the *mutY-mutM* mutation rate estimated by the model and the rates listen in Table E.1.
In the competition experiment, PAO1 is competing with one of the mutator strains for the same substrates. Likewise in the competition model, each of the equations for the three bacteria populations (bacteria, adapted bacteria, and resistant bacteria) is repeated twice. The only parameters differing between PAO1 and each of the four mutators are the mutation rates.

E.2.3 Estimated model parameters

The model parameters shown in Table E.2 are estimated from the growth experiment as described in the previous section. It is found that $\beta = 0$ (p = 0.67), and the model has been adjusted accordingly. The mutation rates are estimated to be 8000 times higher than those listed in Table E.1, which indicates that it is adaptation rather than mutation leading to improved fitness. The result of a simulation of the model is shown in Figure E.4, where the simulations (dotted lines) are plotted together with the observed concentrations (full lines). It is seen that the model captures the dynamics of the growth very well.

E.2.4 Simulation of the competition experiment

The competition model is simulated using the parameter values listed in Table E.2 and the results are plotted in Figure E.5. The Figure shows the total concentration, the wild-type concentration and the mutator concentration as predicted from 10 simulations with the model. The model is seen to capture well the change in total concentration for the competition experiments, even though the model parameters were estimated from the growth experiment and not the competition experiment. The mutator percentage as calculated from the model is shown in

Table E.2: Parameters for growth in LB and LB with ciprofloxacin. The parameters η_1 and η_2 are estimated separately for day 1-2 and for day 3-4 (values separated by comma) for growth in LB. For the non-resistant bacteria growing in LB with ciprofloxacin γ is estimated for day 1 and for day 2-4 (values separated by comma

.

	LB	LB with ciprofloxacin	
		non-resistant	resistant
υ_1	1.56 (0.03)	1.13 (0.02)	1.30 (0.01)
υ_2	0.24 (0.01)	0.22 (0.01)	0.15 (0.02)
η_1 x 10^{10}	5.4 (0.12), 4.4 (0.51)	6.8 (0.85)	4.1 (1.0)
η_2 x 10^{10}	2.2 (0.51), 2.4 (0.82)	3.7 (0.83)	2.2 (0.27)
k	1.30 (2.34)	0.001 (0.0006)	0.137 (0.397)
γ	0.008 (0.001)	0.002 (0.0006), 0.662 (0.063)	-
μ (*mutY-mutM*)	-	-	$2 \cdot 10^{-3}$

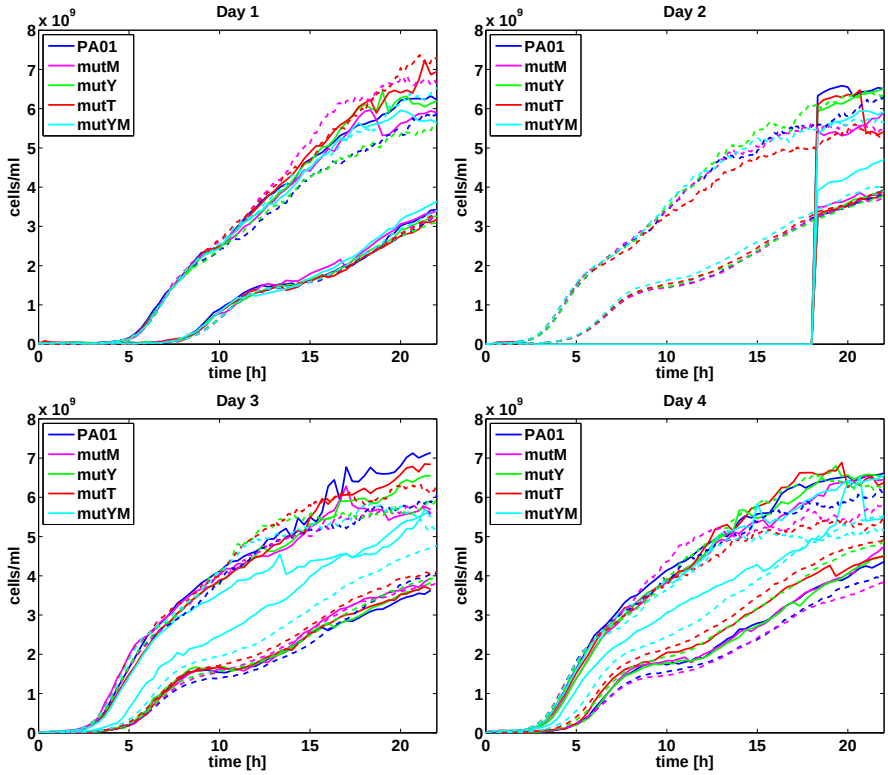
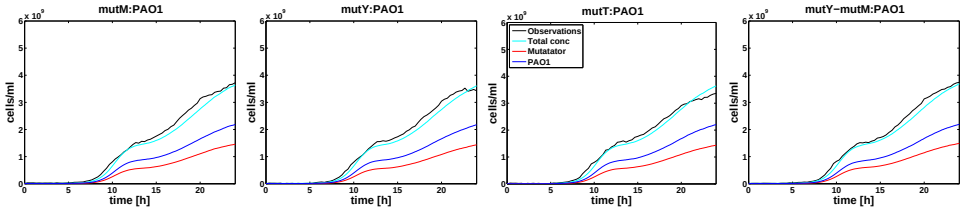
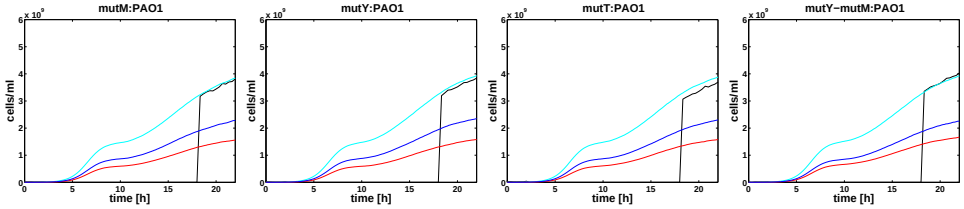


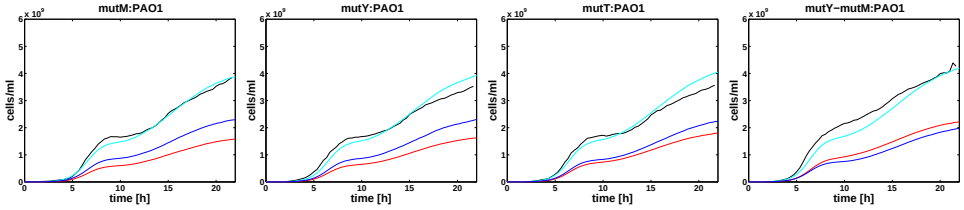
Figure E.4: Simulation results for the growth of each of the five strains in LB (top dotted lines in each subplot), and LB with ciprofloxacin (bottom dotted lines in each subplot) plotted together with data (full line).



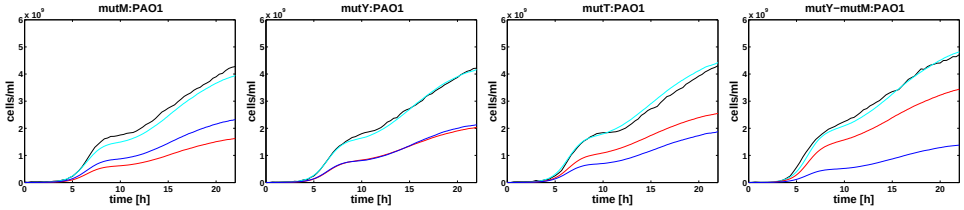
(a) Day 1



(b) Day 2



(c) Day 3



(d) Day 4

Figure E.5: Simulation results and data values for the total bacterial concentration of when in LB with ciprofloxacin. For the simulation model the concentration of PAO1 and each of the four mutator bacteria are also plotted.

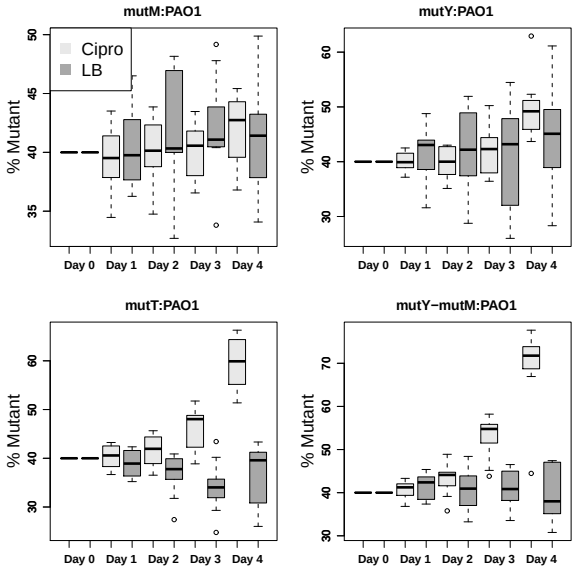


Figure E.6: The percentage of the bacterial population which consists of mutators for the initial population (Day 0) and for each of the following four days of the experiment as estimated from 10 simulations with the competition model.

Figure E.6.

E.3 Discussion

Most information in this study has been gained from the *mutY-mutM* mutant, as the competition experiment showed that the *mutY-mutM* mutator population will take over the whole population, when growing under sub-MIC concentrations of ciprofloxacin. The stochastic model captures the dynamics of the competition between PAO1 and *mutY-mutM* well. From Figure E.3 it is seen that the population consists of a gradually higher concentration of *mutY-mutM*. The same result is obtained from the model as seen in Figure E.6. This is not a coincidence, but is based on a decision on how to estimate the model parameters. Initially we believed that the mutator population would obtain a higher fitness than the wild-type due to a mutation leading to resistance. Therefore the mutation rate was kept fixed on the value estimated by the fluctuation experiment (Figure E.1), and the growth parameters were altered to obtain the best fit to data. This resulted in a higher growth rate than for growth in LB, which seems very unlikely. But more importantly a low mutation rate combined with a high growth rate would result in a sudden increase in the mutator fraction of the population, and not the slow increase observed in the experiment. Furthermore, with a low mutation rate the occurrence of the first resistant bacteria would be a stochastic event, and we would therefore expect a higher variability between the wells in the experiments, as was actually seen. Thus,

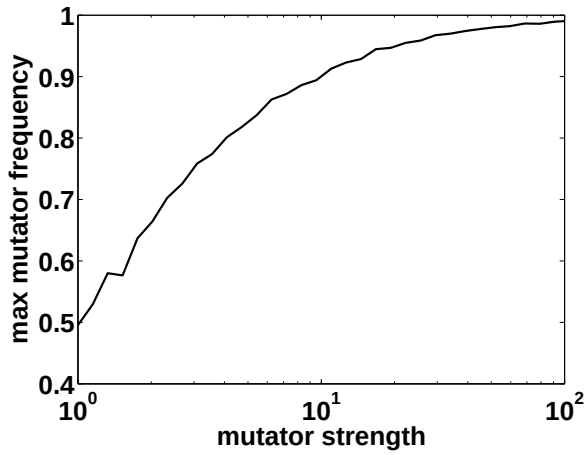


Figure E.7: The equilibrium frequency of a mutator bacteria when growing in competition with PAO1 for different values of the mutator strength. The equilibrium value is the mean of 50 simulations of the model, with the parameters given in Table E.2.

it can be inferred from the model and the data available that it is not a mutation but rather some kind of adaptation that leads to a higher fitness of the mutator when growing in LB with ciprofloxacin. The mutator strains are faster to obtain the necessary adaptations, and thus will outgrow the wild-type strain.
To further examine the relation between mutator strength and competitive growth between a mutator and PAO1, a simulation study was made. A competition experiment was simulated with an initial relation between wild-type and mutator of 1:1, and different values of the mutation strength. The mutation rate of PAO1 was fixed at the value estimated by the model ($2.5 \cdot 10^{-5}$ per cell per generation). The equilibrium frequency of the mutator bacteria depends on the mutator strength as seen in Figure E.7. The equilibrium frequency was reached after 6 to 9 days depending on the mutator strength. The higher the mutation rate of the mutator the faster the equilibrium is reached.

E.4 Conclusion

A mathematical model, based on stochastic differential equations, for competing growth between PAO1 and each of the four mutator strains *mutM*, *mutY*, *mutT*, and *mutY-mutM* has been developed. The model captures the growth dynamics observed in the experiments well. The mutator population is found to gradually take up a larger percentage of the total population when growing in sub-MIC concentrations of ciprofloxacin. Our model indicates that this is due to the faster development of adaptations in the mutator bacteria resulting in a higher fitness. Furthermore, it is found by simulations with the stochastic model that the equilibrium frequency of mutators depend on the mutator strength.

E.5 Bibliography

Bajpai-Dikshit, J., Suresh, A. K., Venkatesh, K. V., 2003. An optimal model for representing the kinetics of growth and product formation by lactobacillus rhamnosus on multiple substrates. J. Biosci. Bioeng. 96 (5), 481–486.

Blázquez, J., Gómez-Gómez, J.-M., Oliver, A., Juan, C., Kapur, V., Martín, S., 2006. PBP3 inhibition elicits adaptive responses in *Pseudomonas aeruginosa*. Mol. Microbiol. 62 (1), 84–99.

Ciofu, O., Riis, B., Pressler, T., Poulsen, H. E., Høiby, N., Jun 2005. Occurrence of hypermutable pseudomonas aeruginosa in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. Antimicrob. Agents Chemother. 49 (6), 2276–2282.

Doshi, P., Rengaswamy, R., Venkatesh, K. V., 1997. Modelling of microbial growth for sequential utilization in a multisubstrate environment. Process Biochem. 32 (8), 643–650.

Ferroni, A., Guillemot, D., Moumile, K., Bernede, C., Bourgeois, M. L., Waernessyckle, S., Descamps, P., Sermet-Gaudelus, I., Lenoir, G., Berche, P., Taddei, F., Aug 2009. Effect of mutator p. aeruginosa on antibiotic resistance acquisition and respiratory function in cystic fibrosis. Pediatr. Pulmonol. 44 (8), 820–825.

Gibson, R. L., Burns, J. L., Ramsey, B. W., 2003. Pathophysiology and management of pulmonary infections in cystic fibrosis. Am. J. Respir. Crit. Care Med. 168 (8), 918–951.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., 2001. Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut. Science 291 (5513), 3082840–2608.

Kompala, D. S., Ramkrishna, D., Tsao, G. T., 1984. Cybernetic modeling of microbial growth on multiple substrates. Biotechnol. Bioeng. 26 (11), 1272–1281.

Kristensen, N. R., Madsen, H., Jørgensen, S. B., 2004. Parameter estimation in stochastic grey-box models. Automatica 40 (40), 225–237.

Lyczak, J. B., Cannon, C. L., Pier, G. B., Apr 2002. Lung infections associated with cystic fibrosis. Clin. Microbiol. Rev. 15 (2), 194–222.

Maciá, M. D., Borrell, N., Segura, M., Gómez, C., Pérez, J. L., Oliver, A., Mar 2006. Efficacy and potential for resistance selection of antipseudomonal treatments in a mouse model of lung infection by hypermutable pseudomonas aeruginosa. Antimicrob. Agents Chemother. 50 (3), 975–983.

Mandsberg, L. F., Ciofu, O., Kirkby, N., Christiansen, L. E., Poulsen, H. E., Hoiby, N., 2009. Antibiotic resistance in *Pseudomonas aeruginosa* strains with increased mutation frequency due to inactivation of the dna oxidative repair system. Antimicrob. Agents. Chemother. 53 (6), 2483–2491.

Mandsberg, L. F., Maciá, M. D., Oliver, A., Philipsen, K. R., Christiansen, L. E., Høiby, N., Ciofu, O., 2010. Development of antibiotic resistance and up-regulation of the antimutator gene pfpi in mutator *P. aeruginosa* due to inactivation of two dna oxidative repair genes (muty mutm), submitted.

Mao, E. F., Lane, L., 1997. Proliferation of mutators in a cell population. J. Bacteriol. 179 (2), 417–422.

Mena, A., Smith, E. E., Burns, J. L., Speert, D. P., Moskowitz, S. M., Perez, J. L., Oliver, A., Dec 2008. Genetic adaptation of pseudomonas aeruginosa to the airways of cystic fibrosis patients is catalyzed by hypermutation. J. Bacteriol. 190 (24), 7910–7917.

Miller, J. H., Suthar, A., 1999. Direct selection for mutators in *Escherichia coli*. J. Bacteriol. 181 (5), 1576–1584.

Montanari, S., Oliver, A., Salerno, P., Mena, A., Bertoni, G., Tümmler, B., Cariani, L., Conese, M., Döring, G., Bragonzi, A., May 2007. Biological cost of hypermutation in pseudomonas aeruginosa strains from patients with cystic fibrosis. Microbiology 153 (Pt 5), 1445–1454.

Oliver, A., Campo, P., Baquero, F., 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science 288 (5469), 1251–1253.

Pal, C., MaciÃ¡, M. D., Oliver, A., Schachar, I., Buckling, A., 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. Nature 450 (7172), 1079–1081.

Philipsen, K., Christiansen, L., Hasman, H., Madsen, H., 2010a. Modelling conjugation with stochastic differential equations. Journal of Theoretical Biology 263 (1), 134–142.

Philipsen, K. R., Christensen, L. E., Mandsberg, L. F., Madsen, H., 2010b. Comparison of calibration curves for the relation between optical density and viable count bacteria data, submitted.

Philipsen, K. R., Christiansen, L. E., Madsen, H., 2010c. Modelling bacterial growth in rich media with a non-parametric extension to a sde based model, submitted.

Philipsen, K. R., Christiansen, L. E., Mandsberg, L. F., Ciofu, O., Madsen, H., Dec 2008. Maximum likelihood based comparison of the specific growth rates for p. aeruginosa and four mutator strains. J. Microbiol. Methods 75 (3), 551–557.

Smith, J. J., Travis, S. M., 1996. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. Cell 85 (2), 229–236.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S., Olson, M. V., 2000. Complete genome sequence of Pseudomonas aeruginosa pa01, an opportunistic pathogen. Nature 406 (6799), 959–964.

Tanaka, M. M., Bergstrom, C. T., Levin, B. R., 2003. The evolution of mutator genes in bacterial populations: The roles of environmental change and timing. Genetics 164 (3), 843–854.

Tenaillon, O., Toupance, B., 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. Genetics 152 (2), 485–493.

Travis, E. R., Travis, J. M. J., 2004. Investigations - mutators in space: The dynamics of high-mutability clones in a two-patch model. Genetics 167 (1), 513–522.

Travis, J. M. J., Travis, E. R., 2002. Mutator dynamics in fluctuating environments. Proc. R. Soc. Lond., B, Biol. Sci. 269 (1491), 591–597.

PAPER

F

Mutators, a way to bypass classical Darwinism [‡]

Abstract

Evolution of antibiotic resistance and other novel bacterial phenotypes continues to happen at such a high rate that it is difficult to explain by classical Darwinian evolution. In Darwinian evolution new mutations must have an advantage compared to the parent type and grow in sufficient numbers to make it plausible that new mutations occur and become established. However, nature also seems to operate an alternative route. Thus, mutators have been suggested as important for increasing bacterial evolution (Oliver et al., 2000; Travis and Travis, 2004; Miller et al., 2002; Denamur and Matic, 2006), due to their high mutation rates and increasing frequencies (Oliver et al., 2000; Mao and Lane, 1997), but their role is not well understood. Here we introduce and model a new hypothesis which can help to explain the advantage of mutator sub-populations and rapid bacterial evolution. Simulation studies show how mutator populations can function as "genetic work stations", where multiple mutations occur and are subsequently transmitted to the parent (wild-type) population by horizontal gene transfer. Thus, mutator populations can be seen as a way to bypass traditional Darwinian evolution.

In classical Darwinism, evolution of new characters is a sequence of single events that each become established in the population before new mutations arise (Darwin, 1859 (reprinted 1998). In bacteria the rapid evolution in antibiotic resistance has been an intriguing example of evolution in real time. The evolution of antibiotic resistance in bacterial populations is also an increasing problem which limits the options for treatment of bacterial infections in humans and animals. Even though it has been possible to identify putative evolutionary pathways when studying TEM genes (Weinreich et al., 2006; Barlow and Hall, 2002), not all the

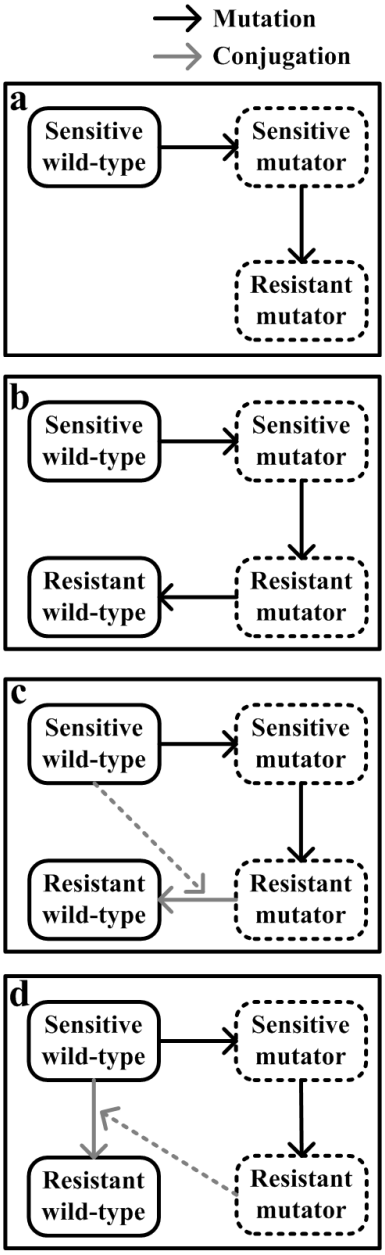
[‡]Submitted as: K. R. Philipsen, L. E. Christiansen, H. Madsen, F. M. Aarestrup, 2009. Mutators, a way to bypass classical Darwinism. Submitted to Nature.

necessary intermediate genes have been observed in nature, nor has it been possible to reach all mutations studied (Barlow and Hall, 2002). For CTX-M enzymes preliminary studies have indicated that double mutants are needed for combined ceftazidime and cefotaxime resistance, while single mutants have reduced susceptibility to one or other of the cephalosporins (Novais et al., 2008).

It is generally accepted that several of the genes conferring antibiotic resistance in clinical bacterial isolates have their origin in the antibiotic-producing organisms. However, the evolution from the *vanA* homolog in the glycopeptides producing organism *Amycolopsis orientalis* to the *vanA* gene located on Tn*1546* would require 480 single mutations each having a selective advantage above the previous in a gene 1,048 bp in length.

In recent decades increased attention has been given to mutator bacteria (Sniegowski and Gerrish, 2000; de Visser, 2002), due to the finding of an increased frequency of mutators in clinical (Oliver et al., 2000) and experimental (Mao and Lane, 1997) bacterial populations. Mutators are present in bacterial populations in natural and clinical environments with frequencies from 10^{-5} (Tanaka et al., 2003), up to 0.01 (Gross and Siegel, 1981; LeClerc et al., 1996) for pathogenic isolates and as high as 0.30 in cystic fibrosis patients (Oliver et al., 2000). Classical theory of evolution concludes that in a stable environment mutation rates will remain low (Tanaka et al., 2003). However, experimental studies (Mao and Lane, 1997; Miller and Suthar, 1999) have shown that the frequency of mutators increases if mutations are necessary to overcome a change in the environment. These findings have been supported by several mathematical models and simulation studies with a single shift (Tenaillon and Toupance, 1999) or constantly fluctuating environments (Travis and Travis, 2004; Tanaka et al., 2003; Travis and Travis, 2002) between two possible states. These studies describe the short-term fate of mutators hitch-hiking to higher frequencies (de Visser, 2002).

The role and fate of mutators is not well understood (de Visser, 2002), but different hypothesis especially related to development of antibiotic resistance have been put forward as illustrated in Figure F.1. These models are generally considered for mutations to all other phenotypes. The basic model (Figure F.1a) suggests that being a mutator is irreversible, and that it is acceptable for a population to have a high frequency of even negative mutations and thereby a lower fitness, because this increases the chance for the population to survive in environments with high exposure to ,e.g., antibiotics. This has been exemplified by *Pseudomonas aeruginosa* infections in cystic fibrosis patients (Oliver et al., 2000). A second model (Figure F.1b) suggests that after obtaining the favourable mutation the mutator genotype will revert to the wild-type allele by mutation, hereby decreasing the mutation rate of the population again (Taddei et al., 1997). Several phylogenetic studies (Denamur and Lecointre, 2000; Brown et al., 2001) argue that instead of reverting to a low mutation rate by mutation, the bacteria can reacquire the mismatched gene



a, A wild-type isolate evolves into a mutator phenotype with an increased ratio of mutations to resistance (Oliver et al., 2000). The increased mutation rate will however, also increase the number of letal mutations.

b, The mutator genotype is obtained by mutation in an anti-mutator gene such as a mismatch repair gene. The resistant mutator strain can re-establish a low mutation rate by reversion (Taddei et al., 1997).

c The mismatch repair gene is lost causing the wild-type bacteria to become a mutator. The mismatch repair gene can be reacquired by the resistant mutator by horizontal gene transfer (Denamur and Lecointre, 2000; Brown et al., 2001).

d, The resistance obtained by the mutator genotype is located on a transferrable plasmid and can therefore be transferred to the wild-type bacteria by horizontal gene transfer [this study].

Figure F.1: Possible pathways of bacterial evolution to antibiotic resistance by mutator populations.

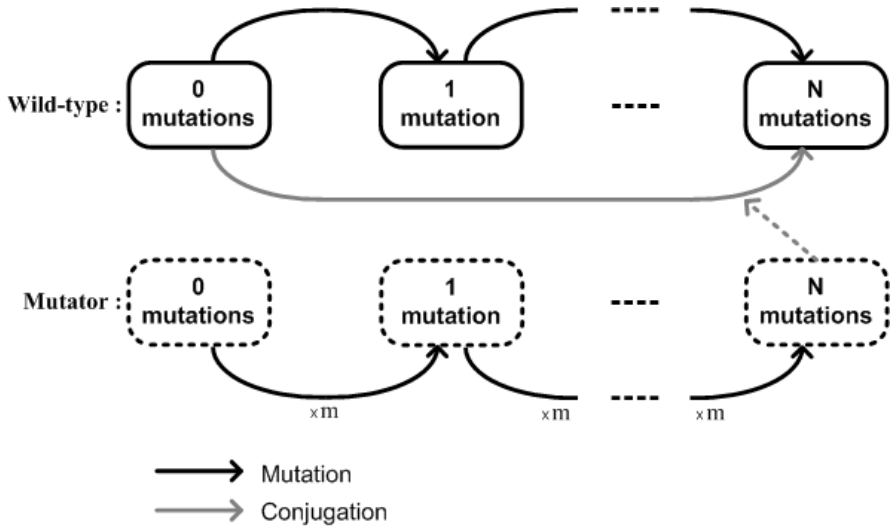


Figure F.2: The wild-type bacteria can acquire mutations via two routes. Either through mutations in the wild-type population, or via the mutator route by receiving a plasmid with mutations from the mutator population.

by horizontal transfer (Figure F.1c). In this study we suggest and discuss another possible evolutionary path (Figure F.1d). Our hypothesis states that the wild-type population will be the lasting population, which obtains the genes conferring resistance by horizontal gene transfer from the mutator population. It will therefore be an advantage for the wild-type population to continuously send out new "genetic working stations" in the form of mutator sub-populations. This would also explain the high frequency of mutators observed in wild-type populations.

We performed a simulation study with up to two consecutive mutations to examine the possibility of acquiring resistance via horizontal transfer from a mutator population (mutator route) as opposed to obtaining resistance from mutations in the wild-type population (wild-type route). The model (Figure F.2) was simulated such that it resembles a possible *in vivo* situation, where bacteria grow exponentially in an environment without antibiotics until a maximum population size is reached and dilution initiated (see Methods for details). The parameter values used for the simulation study were based on existing literature for *Escherichia coli* (see Methods for the values). Since the bacteria were simulated to grow in a competing environment, the mutator population with reduced fitness eventually disappeared from the population as seen in Figure F.3 for one simulation of the model. The simulation was run for five days with a total maximum population of 10^{12} bacteria, initial mutator frequency of 0.01, and a mutator strength, m, of 1,000. The maximum number of single mutations per generation was approximately $1 \cdot 10^{-8} \cdot 9.9 \cdot 10^{11} = 9,900$ for the wild-type population and $1,000 \cdot 1 \cdot 10^{-8} \cdot 8.2 \cdot 10^{9} = 82,000$ for the mutator population. The expected maxi-

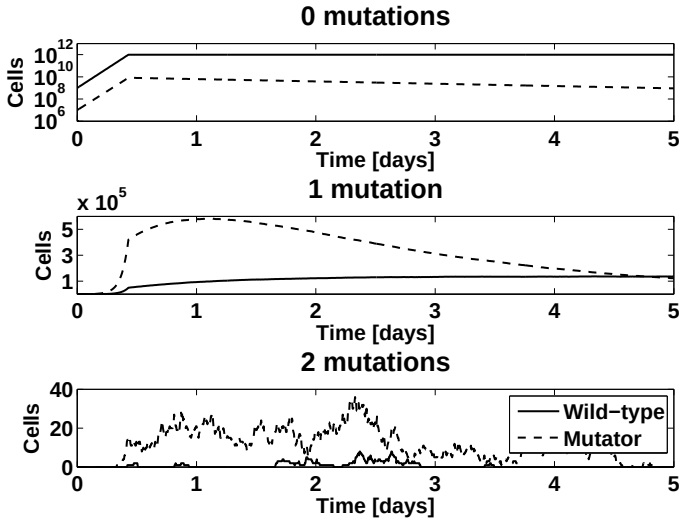


Figure F.3: A simulated time series for the evolution of resistance by mutation and conjugation. The model illustrated in Figure F.2 was simulated with up to two mutations, a mutator strength of 1,000 and a conjugation rate of 0.01. The number of wild-type bacteria with two mutations plotted is the sum of bacteria obtained by two mutations in the wild-type population, and bacteria obtained by conjugation from the mutator population with two mutations. The wild-type bacteria with one mutation were obtained only by mutation in the wild-type population.

mum number of two consecutive mutations per generation was $1 \cdot 10^{-8} \cdot 1.4 \cdot 10^5 = 0.0014$ for the wild-type population and $1{,}000 \cdot 1 \cdot 10^{-8} \cdot 5.8 \cdot 10^5 = 5.8$ for the mutator population. However, due to the lower fitness of the mutator bacteria there was only a short time available to obtain the two mutations and transfer them to the wild-type.

Not surprisingly the wild-type route was dominant when considering just one mutation. The number of wild-type bacteria with one mutation that comes via the wild-type route is approximately a factor of 2 higher than via the mutator route for the simulation in Figure F.3. The number of mutations and conjugation events leading to wild-type bacteria with a given number N of consecutive mutations can be assumed to be Poisson distributed, and the wild-type and mutator routes can therefore be compared by comparing the Poisson parameter for the distributions of mutation and conjugation events. A fit was made to the Poisson distribution obtained for two consecutive mutations from 1,000 simulations. The Poisson parameter was estimated to 0.08 (95% CI = [0.06,0.09]) for the wild-type route and 19.20 (95% CI = [18.93, 19.47]) for the route via the mutator population. Thus, the chance of acquiring two mutations via the mutator subpopulation was significantly higher than via the wild-type route.

Whether the wild-type or mutator route dominates depends on the initial mu-

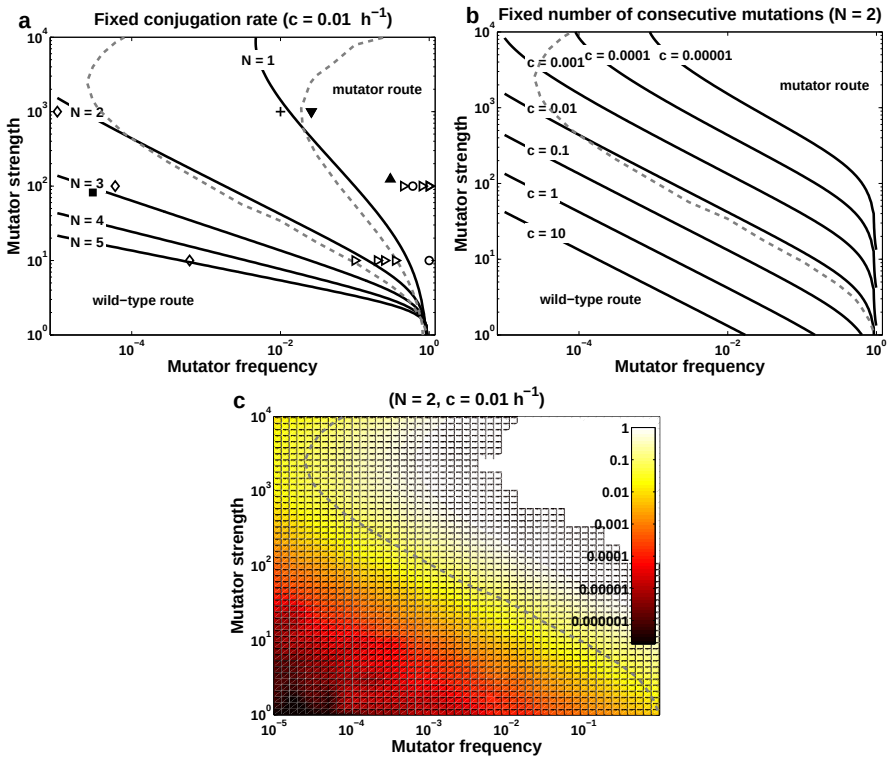


Figure F.4: Comparison of the wild-type route and mutator route. The set of values for which the mutator route and wild-type route are equally dominant were found by calculating the expected number of bacteria resulting from each route (full line) and by comparing the Poisson parameter of the simulation study (dotted line) run for five days. **a**, For one to five mutations and a conjugation rate of $c = 0.01$ h^{-1}. The symbols show the parameter values used in experimental studies (filled symbol) or simulation studies (open symbol) for a given (typically equilibrium) frequency of mutators. The studies referred to are: + (This study Figure F.3), △ (Oliver et al., 2000), ▽ (LeClerc et al., 1996), ▷ (Tenaillon and Toupance, 1999), ◯ (Travis and Travis, 2002), ◇ (Taddei et al., 1997), and □ (Boe et al., 2000). **b**, For two mutations and different conjugation rates. **c**, The probability indicated by colour of at least one wild-type bacteria with two mutations obtained by conjugation from the mutator population for a conjugation rate of 0.01 h^{-1} and simulations for five days. The probability of mutation via the wild-type route is constant (at 0.07), whereas the probability of transfer via the mutator route is seen to increase with increasing mutator strength and mutator frequency.

tator frequency, the mutator strength, the conjugation rate, and the number of consecutive mutations. One approach to relate the two routes is to compare the Poisson distributions obtained from simulations. If the Poisson parameters are identical it means that the two routes are equally likely, as marked with the dotted lines in Figure F.4. However, for low mutation strength the mutator frequency can be assumed to be constant during the five days simulated. In this case a simpli-

fied approach is to consider the expected number of mutation and conjugation events derived from the equilibrium model formulation. If the expected number of events is equal the two routes are equally likely, as marked with the full lines in Figure F.4. The two routes were compared for one to five consecutive mutations (Figure F.4a) and for different values of the conjugation rate (Figure F.4b). The mutator route becomes increasingly important when more mutations are required and for increasing values of the conjugation rate. In Figure F.4a the parameter values from published simulation or experimental studies are marked, hereby placing the parameter values for the two routes in relation to existing literature. The mutator route is seen to be dominant for several plausible values of the mutator frequency and mutator strength. Furthermore, the mutator route becomes increasingly likely for higher mutator strength and mutator frequency as seen from Figure F.4c, where the probability of at least one conjugation event within five days is plotted. An optimum probability is reached for a mutator strength of around 4,000, after which the mutator route becomes less likely, as increased mutator strength is associated with a loss of fitness. The likelihood of transfer via the wild-type route is constant for a given time period.

We have experienced a tremendous evolution of human pathogens during the last decades. Novel pathogens have emerged and old ones have re-emerged. Furthermore, the "new" isolates have often carried novel properties, which are best exemplified by the rapid emergence of antibiotic resistance. Even considering the extreme large number of bacteria colonizing humans and animals this rapid emergence might be hard to explain. Thus, *E. coli* are normally present in numbers of approximately 10^6 to 10^7 per gram faeces. Even assuming an average of one kg of faeces in each human this would only equal around 10^{18} to 10^{19} *E. coli* associated with humans in the world. With an average mutation rate of 10^{-8}, this would only give between 100 to 1000 double mutants per generation. Assuming that these mutants are present in a selective environment and successfully survive several bottle-necks seems almost impossible. Our simulations and model in Figure F.4 with mutators as intermediate "genetic work stations" suggest that the emergence of double, triple, or more consecutive mutations might be more common than so far expected. Thus, evolution of novel features in bacteria might happen more frequently in jumps of several consecutive mutations than we have previously expected. The model might also explain the high level of mutator bacteria observed in wild-type populations, since it will be an evolutionary advantage for a bacterial population to constantly create a number of "genetic work stations" where new genes can be created and tested before being taken up by the mother population.

Whether similar mechanisms of bypassing Darwinian evolution might be present in higher organisms also needs to be proven. However, it is noteworthy that recent studies have indicated that horizontal gene transfer in eukaryotic cells is more important than so far expected (Keeling and Palmer, 2008).

Methods

Procedure for simulations

The model (Figure F.2) is simulated such that it resembles growth with a maximum population size between 10^9 and 10^{12} cells, which is in accordance with bacteria numbers in human faeces (10^{11} cells) and concentrations in infected tissue (10^9 cells/ml) (Miller et al., 2002), and liquid cultures (10^9 cells/ml) (Taddei et al., 1997). The initial cell number is set to 10^9 cells with a frequency of mutator bacteria between 10^{-5} and 0.95. It is assumed that initially no bacteria with mutations are present. When the carrying capacity is reached a dilution is made similar to the growth rate of the sensitive wild-type population. This stochastic model is simulated using a fixed-increment time advance approach (Law and Kelton, 2000) with Poisson-distributed events. A discrete event simulation approach was tested with similar results, but this is computationally infeasible due to the high number of events.

Mutation, conjugation and growth rates

The growth rate v of the wild-type bacteria without any mutations is 0.6931 h^{-1} similar to *E. coli* (Carr et al., 2005). The mutation rate μ is fixed to $1 \cdot 10^{-8}$ per generation (Boe et al., 2000; Tenaillon and Toupance, 1999; Taddei et al., 1997; Travis and Travis, 2002, 2004), from which the mutation rate per hour for each bacteria strain can be calculated as $v\mu$. All mutations are deleterious and cause an additive fitness loss of 0.05 h^{-1} (Travis and Travis, 2004; Tenaillon and Toupance, 1999; Travis and Travis, 2002; Taddei et al., 1997). The mutation rate in the mutator population is increased by a factor m, the mutator strength. Typical values of m lie between 100 and 10000 (Travis and Travis, 2004; Tenaillon and Toupance, 1999; Taddei et al., 1997; Travis and Travis, 2002). It is believed that higher mutator strengths lead to a lower growth rate due to more deleterious mutations. The literature (Tenaillon and Toupance, 1999; Philipsen et al., 2008; Boe et al., 2000; Tanaka et al., 2003) contains different estimates of the disadvantage of being a mutator. Considering that an increased mutator strength gives a decreased growth rate and using an additive fitness cost of 0.002 with a mutator strength of 100 (Tanaka et al., 2003), the growth rate of the mutator can be calculated as $v - (m-1) \cdot 2 \cdot 10^{-5}$ for $m \in [0; 10^4]$. Conjugation rates between 10^{-5} and 10^{-2} per donor per hour (Fernandez-Astorga et al., 1992; Itaya et al., 2006; Lim et al., 2008) and up to 9 per donor per hour (Andrup and Andersen, 1999) have been found in experimental studies. The conjugation rate in the stochastic simulation study is fixed to 0.01 h^{-1}, whereas in the equilibrium model values between 10^{-5} and 10 h^{-1} are used.

F.1 Bibliography

Andrup, L., Andersen, K., 1999. A comparison of the kinetics of plasmid transfer in the conjugation systems encoded by the F plasmid from *Escherichia col*i and plasmid pCF10 from *Enterococcus faecalis*. Microbiology 145 (8), 2001–2009.

Barlow, M., Hall, B. G., 2002. Corrigenda - predicting evolutionary potential: In vitro evolution accurately reproduces natural evolution of the TEM b-lactamase. Genetics 161 (3), 1355–.

Boe, L., Danielsen, M., Knudsen, S., Petersen, J. B., Maymann, J., Jensen, P. R., 2000. The frequency of mutators in populations of *Escherichia coli*. Mutat. Res. 448 (1), 47–55.

Brown, E. W., LeClerc, J. E., Li, B., Payne, W. L., Cebula, T. A., 2001. Phylogenetic evidence for horizontal transfer of mutS alleles among naturally occurring *Escherichia coli* strains. J. Bacteriol. 183 (5), 1631–1644.

Carr, J. F., Gregory, S. T., Dahlberg, A. E., 2005. Severity of the streptomycin resistance and streptomycin dependence phenotypes of ribosomal protein s12 of thermus thermophilus depends on the identity of highly conserved amino acid residues. J. Bacteriol. 187 (10), 3548–3550.

Darwin, C., 1859 (reprinted 1998). On the origin of species by means of natural selection or, the preservation of favoured races in the struggle for life, 1st Edition. Wordsworth.

de Visser, J. A. G. M., 2002. The fate of microbial mutators. Microbiology (Reading) 148 (5), 1247–1252.

Denamur, E., Lecointre, G., 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell 103 (5), 711–721.

Denamur, E., Matic, I., 2006. Evolution of mutation rates in bacteria. Mol. Microbiol. 60 (4), 820–827.

Fernandez-Astorga, A., Muela, A., Cisterna, R., Iriberri, J., Barcina, I., 1992. Biotic and abiotic factors affecting plasmid transfer in *Escherichia coli* strains. Appl. Environ. Microbiol. 58 (1), 392–398.

Gross, M. D., Siegel, E. C., 1981. Incidence of mutator strains in *Escherichia coli* and coliforms in nature. Mutat. Res. 91 (2), 107–110.

Itaya, M., Sakaya, N., Matsunaga, S., Fujita, K., Kaneko, S., 2006. Conjugational transfer kinetics of pLS20 between *Bacillus subtilis* in liquid medium. Biosci. Biotech. Bioch. 70 (3), 740–742.

Keeling, P. J., Palmer, J. D., 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9 (8), 605–618.

Law, A. M., Kelton, W. D., 2000. Simulation modeling and analysis, 3rd Edition. McGraw-Hill.

LeClerc, J. E., Li, B., Payne, W. L., Cebula, T. A., 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science 274 (5290), 1208–1211.

Lim, Y. M., de Groof, A. J. C., Bhattacharjee, M. K., Figurski, D. H., Schon, E. A., 2008. Bacterial conjugation in the cytoplasm of mouse cells. Infect. Immun. 76 (11), 5110–5119.

Mao, E. F., Lane, L., 1997. Proliferation of mutators in a cell population. J. Bacteriol. 179 (2), 417–422.

Miller, J. H., Suthar, A., 1999. Direct selection for mutators in *Escherichia coli*. J. Bacteriol. 181 (5), 1576–1584.

Miller, K., O'Neill, A. J., Chopra, I., 2002. Response of *Escherichia coli* hypermutators to selection pressure with antimicrobial agents from different classes. J. Antimicrob. Chemother. 49 (6), 925–934.

Novais, A., Canton, R., Coque, T. M., Moya, A., Baquero, F., Galan, J. C., 2008. Mutational events in cefotaximase extended-spectrum beta-lactamases of the CTX-M-1 cluster involved in ceftazidime resistance. Antimicrob. Agents Chemother. 52 (7), 2377–2382.

Oliver, A., Campo, P., Baquero, F., 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science 288 (5469), 1251–1253.

Philipsen, K., Christiansen, L., Mandsberg, L., Ciofu, O., Madsen, H., 2008. Maximum likelihood based comparison of the specific growth rates for *P. aeruginosa* and four mutator strains. J. Microbiol. Methods 75 (3), 551–557.

Sniegowski, P. D., Gerrish, P. J., 2000. The evolution of mutation rates: Separating causes from consequences. Bioessays 22 (12), 1057–1066.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H., Godelle, B., 1997. Role of mutator alleles in adaptive evolution. Nature 387 (6634), 700–702.

Tanaka, M. M., Bergstrom, C. T., Levin, B. R., 2003. The evolution of mutator genes in bacterial populations: The roles of environmental change and timing. Genetics 164 (3), 843–854.

Tenaillon, O., Toupance, B., 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. Genetics 152 (2), 485–493.

Travis, E. R., Travis, J. M. J., 2004. Investigations - mutators in space: The dynamics of high-mutability clones in a two-patch model. Genetics 167 (1), 513–522.

Travis, J. M. J., Travis, E. R., 2002. Mutator dynamics in fluctuating environments. P. Roy. Soc. Lond. B Bio. 269 (1491), 591–597.

Weinreich, D. M., Delaney, N. F., DePristo, M. A., Hartl, D. L., 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312 (5770), 111–114.

Acknowledgement This study was supported by the Danish Research Council for Technology and Production Sciences through the grant 274-05-0117 titled "Evolution and adaptation of antimicrobial resistance in bacterial populations".

PAPER

G

Dynamics of spread of intestinal colonization with extended-spectrum beta-lactamases in *E.coli*: a mathematical model ‡

G.1 Introduction

Extended-spectrum beta-lactamases (ESBL) are enzymes that confer resistance to 3rd generation cephalosporins. The prevalence of ESBL producing Enterobacteriaceae has increased drastically since it was first discovered in 1983 in Germany. The increasing prevalence of ESBL is of major concern as it is associated with failure of treatment, prolonged hospitalization and increased costs (Helfand and Bonomo, 2006; Rodriguez-Bano et al., 2006; Collignon and Aarestrup, 2007). In addition, ESBL resistant bacteria often carry co-resistance to other antibiotics, further complicating the treatment of infections (Rodriguez-Bano et al., 2006). In other areas of antibiotic resistance, such as methicillin-resistant *Staphylococcus aureus* (MRSA) (Bootsma et al., 2006) mathematical models have shown to be a strong tool for interpreting the resistance dynamics and investigating possible interventions. In this report we will attempt to develop such a model for ESBL.
The dynamics of ESBL are very complex and differs from MRSA dynamics by the acquisition routes and type of bacteria carrying the resistance. Different types of ESBL have been identified which can be produced by different bacterial species. Until around 2000 mostly ESBL of type TEM/SHV was found in Europe. In Holland the first ESBL of type CTX-M was detected in 1995 (Hall et al., 2002). Before the introduction of CTX-M the ESBLs reported were predominately stemming from *Klebsiella pneumonia* (Paterson et al., 2003). The CTX-M enzyme is

‡Published as: K. R. Philipsen, M. C. J. Bootsma, M. A. Leverstein-van Hall, J. Cohen Stuart, M. J. M. Bonten, 2009. Dynamics of spread of intestinal colonization with extended-spectrum beta-lactamases in *E.coli*: a mathematical model. IMM-Technical report-2009-13.

most frequently associated with *Escherichia coli*, which has caused a switch from *K.pneumoniae* to *E.coli* as the most predominant species among ESBL producers (Markovska et al., 2008). Moreover, ESBL *K. pneumonia* is mainly nosocomial acquired, whereas ESBL-producing *E.coli* is also found in strictly community acquired infections (Cantón et al., 2008; Rodriguez-Bano et al., 2006).
In this study we suggest a new mathematical model to describe the ESBL dynamics. The main objective of the study is to investigate the plausibility of different transmission routes by comparing a mathematical model for the spread of ESBL with data for ESBL prevalence. For instance, the increase in CTX-M ESBL prevalence may be due to horizontal transfer of CTX-M between species (Markovska et al., 2008; Cantón et al., 2008). Other authors have shown the importance of travellers returning from holiday with an ESBL bacteria colonization (Laupland et al., 2008; Pitout et al., 2004). The model will therefore include these routes of horizontal transfer and external acquisition of ESBL together with cross-transfer and mutation. These pathways will be discussed in details in the next section. It has been argued that the use of 3rd and 4th generation cephalosporins in animals has an influence on the increase of ESBL prevalence in humans, as the drug resistance may spread via food or other sources like the ground water (Collignon and Aarestrup, 2007). However, this effect is not included as it would result in an overparametrization of the model due to lack of data.

G.2 Model

Our model describes the spread of intestinal colonization with ESBL. Intestinal colonization is considered, because colonization usually precedes clinical infections (Harris et al., 2007a). Moreover, most colonized patients do not develop overt infections and, hence, it is believed that colonization is more important for the spread than clinical infections (Harris et al., 2007b). The model considers a hospital and its catchment area and consist of two levels of dynamics: 1) The flow of patient; and 2) The flow of bacteria and resistant genes. Each of the two levels of dynamics is described in the following sections.

G.2.1 Flow of patients

Hospitalized patients are divided into high-risk and low-risk wards for the acquisition of resistance. Hospital-based studies have suggested a number of risk factors for the acquisition of ESBL including intensive care unit (ICU) admission, antibiotic usage, and mechanical devices (Laupland et al., 2008; Cantón et al., 2008). We identify high-risk wards (ICU, Surgery, Hematology, and lung diseases) as wards at the University Medical Center Utrecht (UMCU), the Netherlands with a high level of ESBL colonized patients in 2008. This classification of high-risk wards is in agreement with previous studies (Coque et al., 2002).
We hypothesize that frequent readmitted patients have a role in maintaining a high ESBL prevalence in the hospital. Similar to Cooper et al. (2004) we there-

fore allow the probability per unit time to be readmitted to decrease with the time since the most recent hospital discharge. This is modelled by letting discharged patients move first to core groups and from here on into a catchment population with a lower hospitalization rate. Hence, patients discharged from the low- and high-risk wards are separated in two different compartments (core-groups). This flow of patients can be seen in Figure G.1. Additionally, persons can be removed from the community, either because they die or because they move to another municipality. This is implemented in the model by a constant removal rate from the catchment population and the core groups. As soon as a person is removed, it will be replaced by a new person in the catchment population, which is not colonized with resistant bacteria.

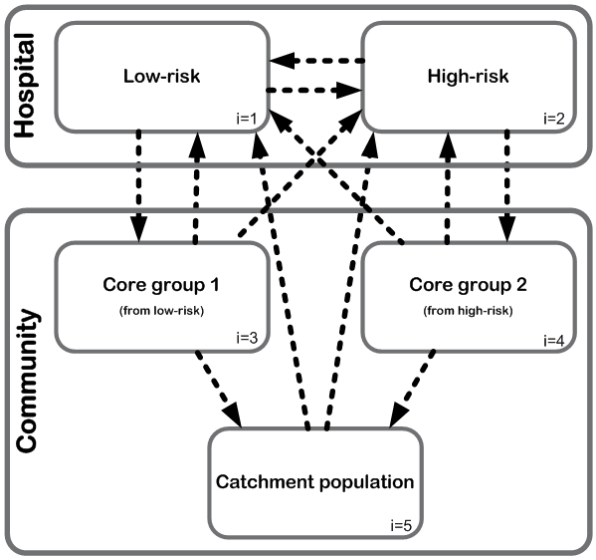
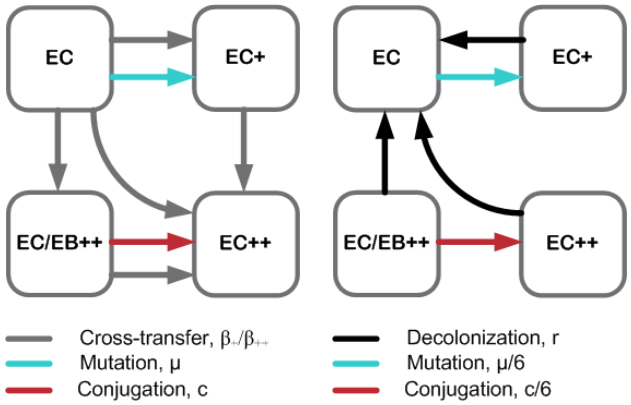


Figure G.1: A sketch of the patient and people flow in the model. The indexes i will be used to refer to the compartments.

G.2.2 Flow of bacteria

Election of colonization states for the model and the routes of transfer between the states is a fine balance between keeping the model simple and including all important states and routes. The extent of the model is further limited by available information about prevalence and rates. ESBL strains are most often found in *E.coli* (EC) and other Enterobacteriaceae (EB) such as *K. Pneumonia* (Romero et al., 2007; Caccamo et al., 2006). For this model we therefore consider EC and another EB type with special focus on the EC population. EC can be ESBL positive of type TEM/SHV (+) or CTX-M (++), which are the dominant types of ESBL. EB is included to be able to incorporate conjugation between species as one of the transfer mechanisms, and therefore the inclusion of EB++ is very interesting

for the model. Assuming each individual carries EC, we distinguish four intestinal colonization states, EC, EC+, EC++ and EC/EB++. The population in each hospital ward and community compartment is divided into these four states (Figure G.2(a) and G.2(b)). The TEM/SHV phenotype can be obtained by cross-transmission,



(a) Hospital (b) Community

Figure G.2: Model describing the transfer of ESBL in *E. coli* (EC) and other Enterobacteriaceae (EB). The model includes the dynamics of the transfer of EC of type TEM/SHV (EC+), EC of type CTX-M (EC++), and EB of type CTX-M (EB++) in the hospital and community. In the hospital the rate of cross-transfer, mutation and conjugation is taken to be three times higher in the high-risk wards, than in the low-risk wards.

and mutation whereas the CTX-M phenotype can be obtained by cross-transmission, conjugation and externally from travellers. The model is constructed on the level of human individuals, and the number of bacteria in each individual is not modelled. Whenever a person acquires resistance, the bacteria is assumed to be present in sufficient numbers for the individual to be able to transfer the resistance. Cross-transmission is dependent on the amount of people in the ward with the specific bacteria species and resistance type. Cross-transmission is therefore modelled with the mass action expression $\beta C/N$, where β is a constant which is specific for a given bacteria species and resistant type, C is the amount of colonized people with the given bacteria species in the ward of interest and N is the total size of the ward. When conjugation happen from a EB++ patient it is no longer registered in the model to carry EB++. To avoid the need of an extra compartment the cross-transfer of EB++ is based on the amount of people colonized with EB++ as well as EC++. Acquisition of ESBL by mutation is independent of the colonization status of other people in the ward. It occurs with a constant rate, μ. Conjugation can occur with a constant rate c, when a patient is colonized with EB++. The rates β, μ and c at which acquisition happens are assumed to be 3 times lower in the low-risk wards as compared to the high-risk wards. Cross-transmission is disregarded in the community, whereas mutation and conjugation occur with half the rate in the community as compared to the low-risk ward. After the year 2000

there is an extra inflow of EC++ and EB++ to the catchment population from travellers carrying the strains home from holiday. This inflow is assumed to occur to the core group 2 for EC++ and EB++ and to the catchment population for EC++ (Pitout et al., 2004). In the community colonization is lost with the rate r (recovery rate). The decolonization rate is the same for all community groups. Whenever possible the model parameters are found in the literature and the remaining parameters will be estimated. A description of the parameter estimation is given in Section G.3.2.
The model is simulated as a discrete stochastic model in R (R Development Core Team, 2009) using the fixed-increment time advance method. For each day the following steps are carried out

- transfer between bacteria colonization states within each hospital ward or community group.
- movement of people within the hospital and community as well as hospital admittance and discharge.
- inflow of resistant strains from travellers.

The two first steps are computed by sampling from a Multinomial distribution, as more events can happen to one population during one time interval. The last step is computed by sampling from a Poisson distribution.

G.3 Parameter estimation

G.3.1 Patient flow

Data from the UMCU from 2005 to 2008 with time of admission, discharge and movement within the hospital, including ward specification, is used to estimate the parameters for the patient flow.

Survival analysis

Part of the data for the time between two hospital admissions is censored, and survival analysis is therefore used to calculate the time to readmission. In this context, censored data means that only the time between discharge and the end time of the available data set is know. Thus, only a minimum time between readmissions is known, but there is no observed readmission. For other patients two subsequent admissions are registered in the data set. Thus, for these patients the actual readmission time is known. These readmission times are also called un-censored.
People discharged from the hospital can be readmitted from either the core-group or catchment population as sketched in Figure G.3. The survival function $S_i(t)$, for readmittance after discharge from the hospital compartment i is calculated as

$$S_i(t) = P_i(T > t) = P_i(T = \infty) + P_i(\infty > T > t) \tag{G.1}$$

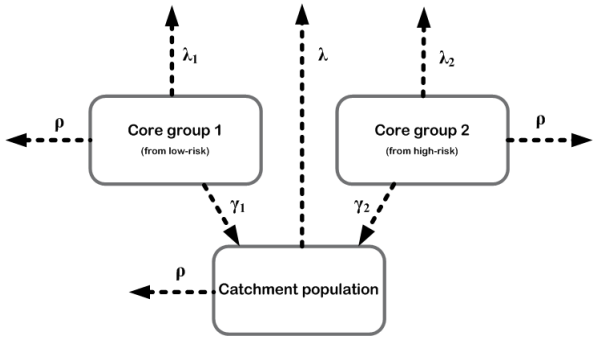


Figure G.3: A sketch of the simplified patient flow from with the survival function for readmission time (Equation (G.4)) is calculated. λ_1, λ_2 and λ are readmission rates and γ_1 and γ_2 are the rates by which patients are moved to the catchment population. The removal rate, ρ, is fixed to one over the mean time of stay in the same municipality, which in the Netherlands is 21.6 years.

where the first term corresponds to patients who are never readmitted, i.e., removal from the extramural population, and the second term to a readmission at least a time t after the previous admission.

$$P_i(T=\infty) = \int_0^\infty \rho \exp(-(\rho+\lambda_i+\gamma_i)\tau)d\tau + \int_0^\infty \gamma_i \exp(-(\rho+\lambda_i+\gamma_i)\tau)\rho\exp(-(\rho+\lambda)\tau)d\tau \quad \text{(G.2)}$$

and

$$P_i(T>\infty>t) = \int_t^\infty \lambda_i \exp(-(\rho+\lambda_i+\gamma_i)\tau)d\tau + \int_t^\infty \left(\int_0^\tau \gamma_i \exp(-(\lambda_i+\gamma_i+\rho)\tau')\lambda\exp(-(\lambda+\rho)(\tau-\tau'))d\tau'\right)d\tau \quad \text{(G.3)}$$

By solving the integrals the survival function is found to be

$$S_i(t) = \frac{\rho}{\rho+\lambda_i+\gamma_i} + \frac{\rho\gamma_i}{(\rho+\lambda)(\rho+\lambda_i+\gamma_i)} + \frac{\lambda_i}{\rho+\lambda_i+\gamma_i}\exp(-(\rho+\lambda_i+\gamma_i)t) - \frac{\lambda\gamma_i}{(\lambda_i+\gamma_i-\lambda)}\frac{1}{(\rho+\lambda_i+\gamma_i)}\exp(-(\rho+\lambda_i+\gamma_i)t) + \frac{\lambda\gamma_i}{(\lambda_i+\gamma_i-\lambda)}\frac{1}{\lambda+\rho}\exp(-(\lambda+\rho)t)\,, \quad \text{(G.4)}$$

where the readmission rate λ_i and the rate of movement to the catchment population γ_i are different for patients discharged from the low-risk (i=1) and high-risk (i=2) wards. The readmission rate from the catchment population, λ is the same, independently of which ward a patient was discharged from at last hospitalization. The removal rate, ρ, is assumed to be the same in the whole community and

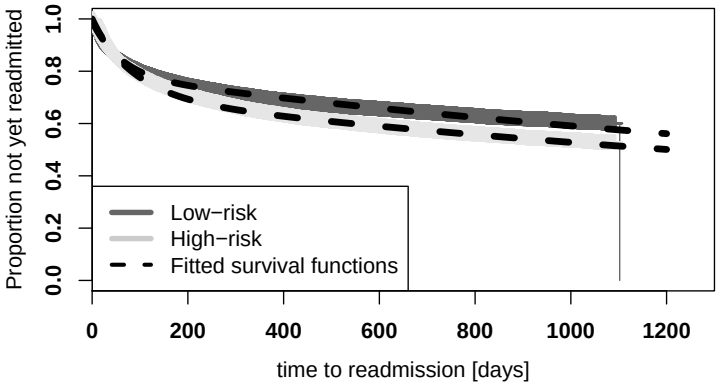


Figure G.4: Fit of the survival function for people discharged from high-risk and low-risk wards at University Medical Centre Utrecht. The data used is from 2005 to 2008. High-risk wards are ICU, Surgery, Hematology and lung diseases, which are identified as those with a high risk of colonization with ESBL carrying bacteria.

is kept fixed in the estimation. A person is removed from the catchment population or core groups after a mean of 21.6 year, which corresponds to the mean time that persons stay in the same municipality in the Netherlands. This is considered to be a good estimate of the time period in which people might be readmitted to the same hospital. The other parameters of the survival function are determined by Maximum Likelihood estimation, where the log-likelihood function is

$$\log(L) = \sum_{\text{i} \in \text{ uncensored data}} \log\left(\frac{f_i(t_i)}{S_i(t_i)}\right) + \sum_{\text{i} \in \text{ all data}} \log(S_i(t_i)), \tag{G.5}$$

and $f(t) = -dS(t)/dt$. The likelihood function is optimized in `R` using the `optim` function (R Development Core Team, 2009).

Results

The suggested survival function gives a good fit to the data for readmission as seen in Figure G.4. It should be noted that the survival function does not describe *to* which ward the patients are readmitted. It only states what the readmission time is when discharged *from* a specific ward. Based on the UMCU patient data we find the percentage of patients discharged from the low- or high-risk wards, which are readmitted to either of these wards, and from here the readmission rates of the core-groups are calculated.
The readmission rate from the catchment population to each hospital ward and the size of the catchment population are estimated, such that the size of the hospital wards are in agrement with the UMCU data. From the UMCU data the length

Table G.1: Parameters for patient flow calculated from UMCU data with a 95% confidence interval (CI.

Parameter	Value (95% CI)
Mean length of stay:	
Low-risk	5.32 (5.27 - 5.36) days
High-risk	6.68 (6.61 - 6.76) days
Mean time to admission:	
From Core-group 1	247.57 (239.53 - 255.62) days
From Core-group 2	293.71 (282.29 - 305.23) days
From Catchment population	8.97 (8.63 - 9.30) years
Mean time before moving to catchment group:	
Core-group 1 to catchment:	73.88 (70.56 - 77.20) days
Core-group 2 to catchment:	130.55 (122.94 - 138.16) days

Table G.2: The table shows data values for the percentage of patients that moved within the UMCU, were discharged, or died at the UMCU between 2005 and 2008. The table also contain the percentage of people admitted to each hospital ward from the community.

To: / From:	Low-risk	High-risk	Discharged	Dead
Low-risk	-	13%	86%	1%
High-risk	24%	-	73%	3%
Core-group 1	87%	13%		
Core-group 2	23%	77%		
Catchment population	72%	28%		

of stay in each ward, the transfer between wards, and the fraction of patients which dies at the hospital are also deduced. As soon as a person is removed from the population a new person, colonized with non ESBL *E.coli*, is assumed to appear in the catchment population. All parameters for the patient flow are given in Table G.1 and Table G.2.

G.3.2 Bacteria flow

The duration of colonization after discharge from the hospital have recently been examined in a study by Apisarnthanarak et al. (2008). They found a median duration of outpatient colonization with ESBL of 98 days, i.e. a mean duration of 141 days. This length of colonization will be used for all ESBL producing bacteria in the model. The inflow of resistance from travellers is initiated in year 2000. The inflow of EC++ and EB++ to core group 2 is fixed to 0.35 per year, and the inflow of EC++ to the catchment population is fixed to 5 per year (Pitout et al., 2004). The unknown parameters for the flow of bacteria are determined in three steps. First in Section G.3.2 three values of the cross-transfer rate of EC+ are determined such

that the basic reproduction number equals 0.50, 0.75 or 1.00. Secondly for each of the estimations of the cross-transfer rate, the mutation rate for EC+ is estimated based on the prevalence of ESBL in The Netherlands before the large increase in ESBL caused by CTX-M. The prevalence of EC+ is assumed to have reached an equilibrium level before the introduction of CTX-M. The used prevalence data is from the study of Stobberingh et al. (1999) which found that the ESBL prevalence at Dutch hospitals in 1997 was 0.35% for EC+. Finally, the parameters for the transfer of EC++ are estimated using nosocomial ESBL prevalence data deduced from the Dutch EARSS data (2000-2008). The EARSS prevalence is based on aggregated data for all of the Netherlands. The parameter estimation of the model can therefore be based on the mean of several simulations. Thus, estimation based on the ESBL prevalence is made by minimizing the least squares (LS) value of the difference between the simulated mean yearly ESBL prevalence based on up to 50 simulations and the data. Due to the stochasticity of the model the LS value will be noise even when taking the mean over several simulations. Therefore the LS estimation will be carried out using a simultaneous perturbation stochastic approximation (SPSA) algorithm explained in Section G.3.2.

Reproduction Number

The basic reproduction number, R_0, is the mean number of secondary colonizations that one colonized individual will cause before it get decolonized. This number is often used in epidemiology, as it can help determine whether an infection (or in this case colonization) will spread in a population. If $R_0 < 1$ the infection will die out, and if $R_0 > 1$ the infection can spread in a population. Alternatively the single readmission number, R_A, can be calculated, which is the mean number of secondary colonizations that one colonized individual will cause during one hospital admissions. We calculate R_0 and R_A for a simplified situation where a person can be either susceptible (i.e. in the category EC) or colonized with EC+. In this way the initial colonization of EC+ bacteria can be studied. We only calculate the reproduction number for cross-transmission, hence the only way of acquiring EC+ is by cross transfer in the hospital with rate β in the high-risk wards and $\beta/3$ in the low-risk wards. The colonization can be lost in the community with the rate r.
R_0 and R_A are found as the largest eigenvalue for the next-generation matrix $\boldsymbol{K}^0$ and $\boldsymbol{K}^A$, respectively (Diekmann and Heesterbeek, 2000). Each element of $\boldsymbol{K}$, k_{ij} is the expected number of new colonized people in compartment i caused by one person colonized in compartment j either during one hospitalization (R_A) or during the whole duration of the colonization (R_0). The expected number of new colonized people in compartment i is given by the transfer rate β multiplied with the time, u, spend in this compartment before discharge from the hospital or loss of colonization. The movement between compartments can be considered as a Markov jump process containing each of the transient states and a compartment representing the absorbing state, i.e. discharge of a patient or loss of colonization.

The process has an intensity matrix of the form

$$\Lambda = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 0 \end{bmatrix} \tag{G.6}$$

where $\boldsymbol{T}$ describes the rate of movement between the transient states and $\boldsymbol{t}$ contains the rates by which exit to the absorbing state takes place. The probability of going from state i to j is written as p_{ij}, for example the probability of going from core group 1 to the low-risk ward is

$$p_{31} = \lambda_{31}/(r + \gamma_1 + \rho + \lambda_{31}) , \tag{G.7}$$

where λ_{31} is the readmission rate from core-group 1 to the low-risk ward, r is the decolonization rate, γ_1 is the rate of transfer from core group 1 to the catchment population, and ρ is the removal rate of persons from the population. If we apply this notation to the model for the patient flow shown in Figure G.1, we can write down the complete $\boldsymbol{T}$ matrix for the case where discharges from the hospital are considered as the absorption state

$$\boldsymbol{T}^A = \begin{bmatrix} -1/T_1 & p_{12}/T_1 & p_{13}/T_1 \\ p_{21}/T_2 & -1/T_2 & p_{23}/T_2 \\ p_{31}/T_3 & p_{32}/T_3 & -1/T_3 \end{bmatrix} \tag{G.8}$$

or when loss of colonization is considered as the absorption state

$$\boldsymbol{T}^0 = \begin{bmatrix} -1/T_1 & p_{12}/T_1 & p_{13}/T_1 & p_{14}/T_1 & 0 & 0 \\ p_{21}/T_2 & -1/T_2 & p_{23}/T_2 & 0 & p_{25}/T_2 & 0 \\ p_{31}/T_3 & p_{32}/T_3 & -1/T_3 & p_{34}/T_3 & 0 & 0 \\ p_{41}/T_4 & 0 & p_{43}/T_4 & -1/T_4 & 0 & p_{46}/T_4 \\ 0 & p_{52}/T_5 & 0 & 0 & -1/T_5 & p_{56}/T_5 \\ p_{61}/T_6 & p_{62}/T_6 & p_{63}/T_6 & 0 & 0 & -1/T_6 \end{bmatrix} \tag{G.9}$$

In both cases $1/T_i$ is the mean length of stay in the given compartment i, as shown in Figure G.1. The time until absorption τ is said to have a phase type distribution $PH(\pi, \boldsymbol{T})$, where π is the initial distribution. It can be shown that if $\boldsymbol{U} = (-\boldsymbol{T})^{-1}$, then each element u_{ij} of the matrix $\boldsymbol{U}$ is the expected time spent in state j given initiation in state i prior to absorption. In this way each element in the next generation matrix $\boldsymbol{K}$ can be found as

$$\boldsymbol{K} = (-\boldsymbol{T})^{-1}_{ij} \beta_j = u_{ij} \beta_j , \tag{G.10}$$

where $\beta_4 = \beta_5 = \beta_6 = 0$, as there is no cross-transfer in the community. R_0 and R_A can then be found as the largest eigenvalue of $\boldsymbol{K}^0$ and $\boldsymbol{K}^A$, respectively

Stochastic approximation

The bacteria transfer parameters will be determined by LS estimation. Due to the stochasticity of the model the object function will be noisy even when taking

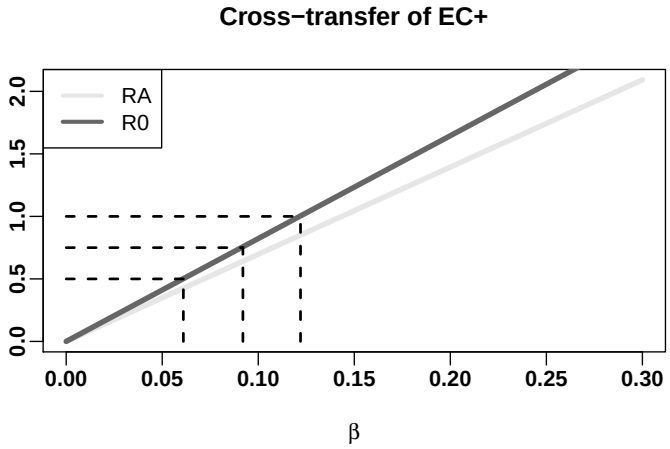
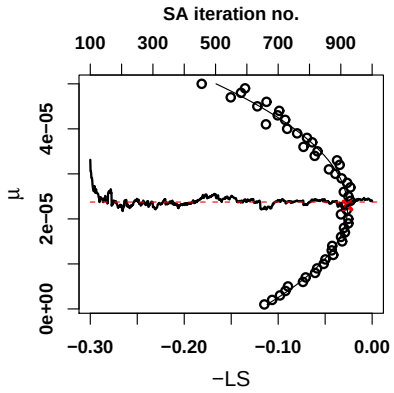


Figure G.5: R_0 and R_A for different values of the cross-transfer rate in the high-risk wards. R_0 equals 1.00, 0.75 and 0.50 for β_+ values of 0.122, 0.092, and 0.061, respectively. The decolonization rate is fixed to $1/141$ days^{-1}.

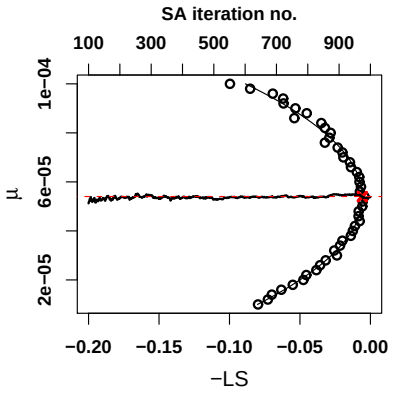
the mean over several simulations. We therefore use a stochastic approximation method to find the minimum object function. The computation of each object function is very time consuming as the model has to be simulated repeatedly for several years to obtain a mean value for the prevalence which can be compared with the observations. Therefore it is desirable to keep the number of evaluations of the object function low. When more than one parameter has to be estimated simultaneously as is the case for the transfer of EC++, the Simultaneous Perturbation Stochastic Approximation (SPSA) method can be used. The SPSA method uses only two measurements of the object function to approximate the gradient, whereas the Finite Difference Stochastic Approximation (FDSA) method uses two measurements per parameter. When only one parameter needs to be estimated the SPSA method simply reduces to the FDSA method. For both methods the aim is to find the set of parameters, for which the gradient of the object function equals zero. A good description of the SPSA algorithm and its implementation can be found in Spall (1998).

Results

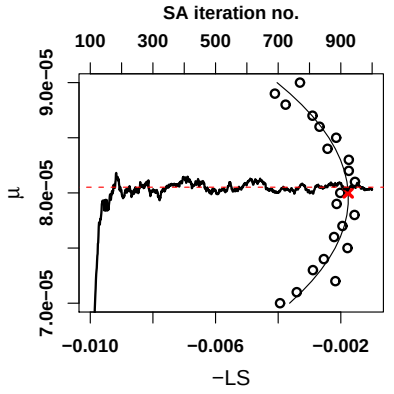
The basic reproduction number, R_0 and the single admission reproduction number, R_A are computed in R (R Development Core Team, 2009) for difference values of the cross-transmission rate, β. The result is plotted in Figure G.5. The cross-transmission rate in the high-risk wards is found to be 0.122, 0.092, and 0.061 in order to give a basic reproduction number of 1.00, 0.75, and 0.50, respectively. The unit for the rates is day^{-1}. For the three estimates of the cross-transmission



(a) $\beta_+ = 0.122$



(b) $\beta_+ = 0.092$



(c) $\beta_+ = 0.061$

Figure G.6: Estimate of the mutation rate, μ using a stochastic approximation method to find the parameter region giving the lowest least squares (solid black line). Subsequently, the mean of 50 evaluation of the LS for different rates in this region is computed (open circles), and the minimum is found by fitting a second order polynomial (solid grey line) to these values and finding the minimum. The mutation rate is found to be $2.3{\cdot}10^{-5}, 5.4{\cdot}10^{-5}$, and $8{\cdot}10^{-5}$ (dotted horizontal line) for corresponding β_+ values of 0.122, 0.092, and 0.061, respectively.

rate, the mutation rate shown in Figure G.6 is found by stochastic approximation and evaluation of the least squares value. The mutation rates are found to be $2.3 \cdot 10^{-5}, 5.4 \cdot 10^{-5}$, and $8 \cdot 10^{-5}$ for corresponding β values of 0.122, 0.092, and 0.061, respectively.

The cross-transfer rate for EC++ and EB++, and the conjugation rate is found by evaluation the LS value for different values of the cross-transfer rate, β_{++}, and conjugation rate, c, as seen in Figure G.7. The intension was to use SPSA to estimate

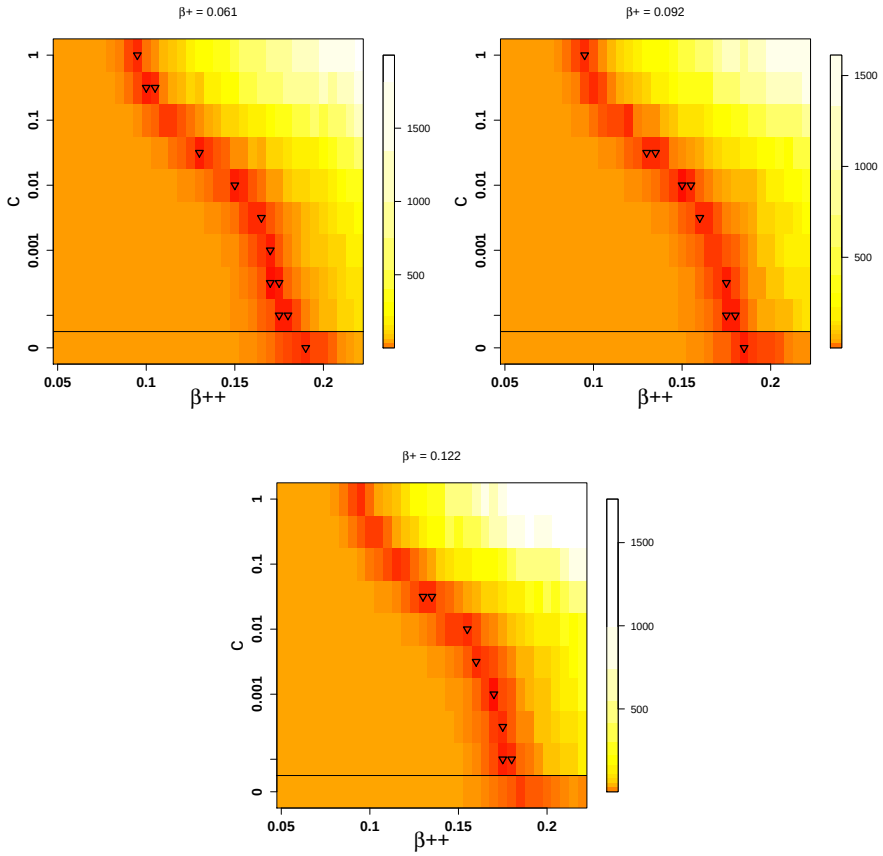


Figure G.7: Least square values (colors) for different combination of the cross transfer rate, β_{++}, of EC++ and EB++, and the conjugation rate, c. Each estimated LS value is the mean of 10 repetitions for which the prevalence is calculated as the mean of 50 simulations of the model. The value of c below the line is 0, whereas it about the line are from 10^{-4} to 1. The triangles marks the least square values below 3.

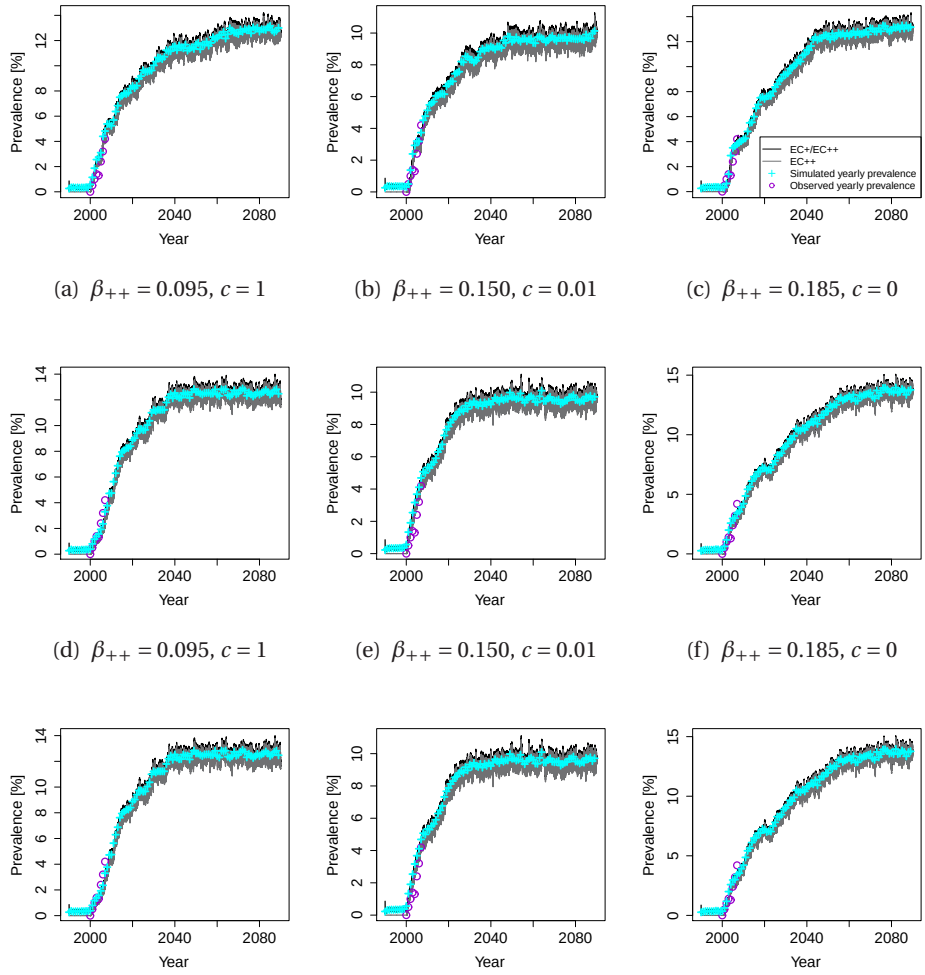
the parameters, but it has not been possible to find a proper implementation of the coefficients for the SPSA method, and it was therefore chosen to compute the LS value for a span of model parameter values.
It is seen to be difficult to separate the conjugation rate and the cross-transfer rate. This is caused by the low amount of data available for the estimation, and the implementation of the cross-transfer to the EC/EB++ colonization state in the model. According to the least squares estimation, the conjugation rate can take on values from 0 to 1 per day, whereas the cross transfer can take on values between 0.095 and 0.19 per day.

G.4 Investigating the spread of resistance

The change over time in the total EC+ and EC++ prevalence and the EC++ prevalence is plotted in Figure G.8, for the three estimates of β_+ and μ, and for each of these three pairs of β_{++} and c values: $\beta_{++} = 0.095$ and $c = 1$, $\beta_{++} = 0.150$ and $c = 0.01$, and $\beta_{++} = 0.185$ and $c = 0$. The model predicts that the total point-prevalence will increase to values between 9.6% and 13.7% depending on the parameter values used. The prevalence in each of the hospital wards from 1990 to 2009 can be seen in Figure G.9 for one combination of model parameters. The equilibrium prevalence for each of these wards are: low-risk ward 5.0%, high-risk ward 18.1%, core group 1 4.1%, core group 2 14.0%, and catchment population 0.5%. The prevalence in the high-risk ward is thus 3.6 times higher than in the low-risk ward, and similarly the prevalence in core group 2 is 3.4 times higher than the prevalence in core group 1. This is caused by the patient flow, as patients discharged from the high-risk ward is most frequently (77%) also readmitted to the high-risk ward. The last subplot in Figure G.9 show a close-up of the total estimated yearly mean prevalence for EC+ and EC++ in the hospital until year 2010.
With the parameters used for Figure G.9 the mean number of patients becoming colonized during one day with either EC+ or EC++ in the high-risk ward in the year 2009 has been calculated. The mean number of occupied beds in the high-risk wards is 183. Out of these patients 0.04 and 0.01 patients per day will be colonized with EC+ due to cross-transfer and mutation, respectively; and 1.92 and 0.16 patients will be colonized with EC++ due to cross-transfer and conjugation, respectively.
The relative importance of each of the three transfer mechanisms: cross-transfer, conjugation, and mutation for the high-risk ward in the year 2009 is plotted in Figure G.4. The model predicts that most transfers of EC++ and EC+ will happen due to cross-transfer. Furthermore, the model predicts that a minimum of 57% patients will acquire EC++ due to cross-transfer (for $c = 1$) in the high-risk ward. However, it can not be ruled out that cross-transfer is the only transfer mechanism for EC++ in the hospital, i.e. that $c = 0$.

G.5 Conclusion and outlook

In this study a mathematical model for the spread of ESBL resistant *E.coli* among patients in a hospital and the surrounding catchment population has been introduced and used to described prevalence data from the Netherlands. Several statistical methods have been applied to estimate the model parameters. The patient flow data was studied by survival analysis. This enabled us to get an estimate for the time to readmission when discharged from either the low-risk or high-risk hospital wards. It is hypothesized that readmission plays a role for the spread of resistant bacteria. The high prevalence of EC+ and EC++ colonized patients in the core group with patients discharged from the high-risk ward, indicates that



(a) $\beta_{++} = 0.095$, $c = 1$ (b) $\beta_{++} = 0.150$, $c = 0.01$ (c) $\beta_{++} = 0.185$, $c = 0$

(d) $\beta_{++} = 0.095$, $c = 1$ (e) $\beta_{++} = 0.150$, $c = 0.01$ (f) $\beta_{++} = 0.185$, $c = 0$

(g) $\beta_{++} = 0.095$, $c = 1$ (h) $\beta_{++} = 0.150$, $c = 0.01$ (i) $\beta_{++} = 0.185$, $c = 0$

Figure G.8: The mean of 50 simulations of the model for different values of the model parameters. In the top panel $\beta_+ = 0.061$ and $\mu = 8 \cdot 10^{-5}$, in the middle panel $\beta_+ = 0.092$ and $\mu = 5.4 \cdot 10^{-5}$, and in the bottom panel $\beta_+ = 0.122$ and $\mu = 2.3 \cdot 10^{-5}$. A total prevalence of EC and EC++ between 9.6% and 13.7% is predicted to be reached in year 2030 or later.

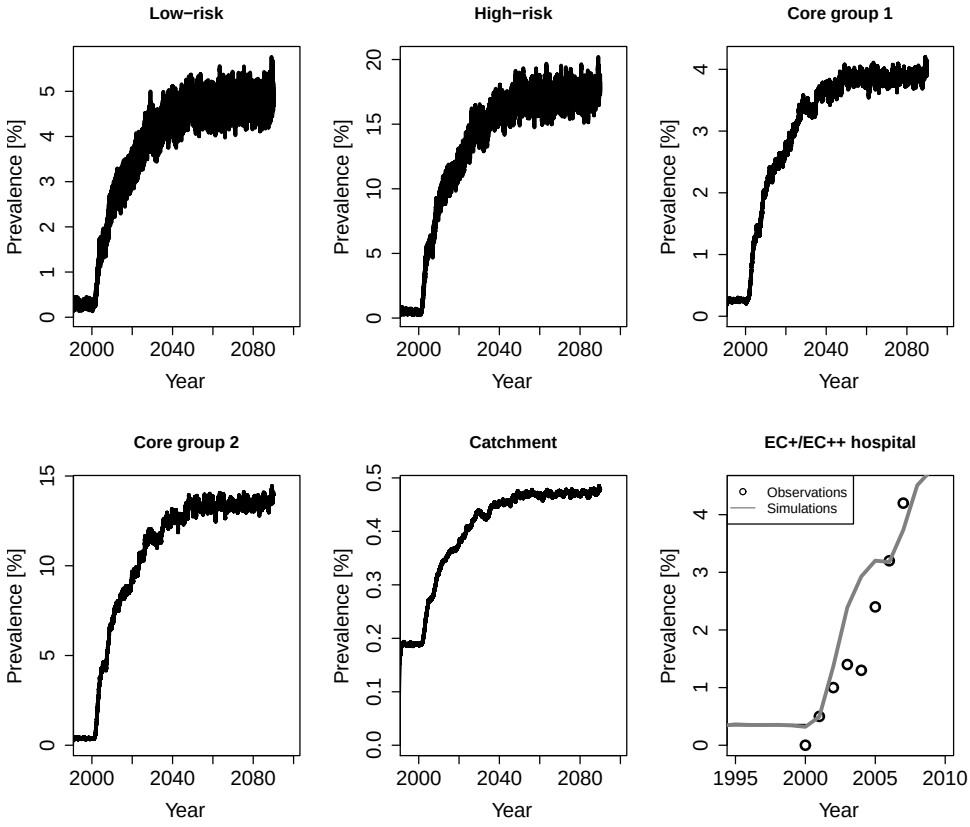


Figure G.9: The total mean prevalence of EC+ and EC++ in each hospital ward and community compartments from 1990 to 2009 for 50 simulations. The parameters used for the simulation was: $\beta_+ = 0.061$, $\mu = 8 \cdot 10^{-5}$, $\beta_{++} = 0.150$, and $c = 0.01$. The plot in the lower right corner show the observed and simulated prevalence EC+ and EC++ prevalence in the hospital.

especially patients readmitted to the high-risk wards contribute to the increasing prevalence. It could be interesting in a future simulation study to further investigate the importance of readmission and the effect of different interventions on the prevalence.

There are several theories with regards to the spread of ESBL resistant bacteria, but the actual prevalence data is very sparse. Based on the available data we have developed an adequate model that can explain the increase in prevalence from year 2000. It has not been possible to separate the effect from conjugation and cross-transfer on the ESBL prevalence of type CTX-M, as several combination of the cross-transfer rate and conjugation rate give a good fit to data. However, the model predicts that a minimum of 57% of the acquisition of EC++ colonization is due to cross transfer.

The transfer rates for each hospital and community compartments are all related

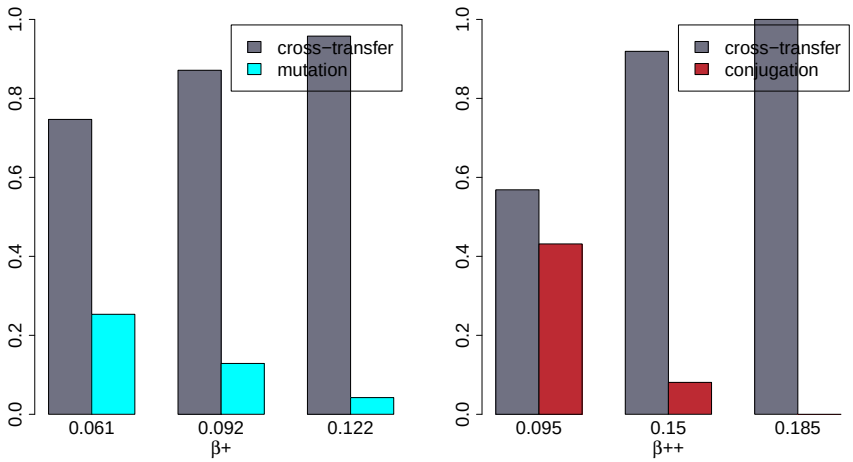


Figure G.10: The relative importance of each of the three transfer mechanisms: cross-transfer, conjugation and mutation, for the transfer of EC+ ($\beta_{++} = 0.150$, $c = 0.01$) and EC++ ($\beta_{+} = 0.061$, $\mu = 8 \cdot 10^{-5}$) as measured in the high-risk ward in year 2009 from 50 simulations of the model.

by a ratio fixed in the model. Due to the high prevalence of resistant bacteria in the high-risk ward, this ward is a central element for estimation of the model parameters. It could therefore be interesting to look at the transfer going on inside the high-risk wards alone. A surveillance study, where the colonization status of all patient in one or two hospital high-risk hospital wards are followed over a couple of month, could be an idea for a better understanding of the transfer mechanisms. The mean duration of colonization with EC+, EC++ and EB++ after discharge from the hospital has in this study been fixed to 141 days. Whether patients readmitted to the hospital are colonized with resistance bacteria is among other things dependent on the duration of colonization. It would therefore be interesting to investigate the effect of increased or decreased length of colonization by simulation studies. Furthermore the model could be improved, if data from colonization studies of each of the colonization states EC+, EC++ and EB++ were available.

G.6 Bibliography

Apisarnthanarak, A., Bailey, T. C., Fraser, V. J., 2008. Duration of stool colonization in patients infected with Extendend-Spectrum β-Lactamase-preducing Escherichia coli and Klebsiella pneumoniae. Clin. Infect. Dis. 46, 1322–1323.

Bootsma, M. C. J., Diekmann, O., Bonten, M. J. M., 2006. Controlling methicillin-resistant staphylococcus aureus: quantifying the effects of interventions and rapid diagnostic testing. Proc. Natl. Acad. Sci. U. S. A. 103 (14), 5620–5625.

Caccamo, M., Perilli, M., Celenza, G., Bonfiglio, G., Tempera, G., Amicosante, G., 2006. Occurrence of extended spectrum β-lactamases among isolates of enterobacteriaceae from urinary tract infections in southern italy. Microb. Drug Resist. 12 (4), 257–264.

Cantón, R., Novais, A., Valverde, A., Machado, E., Peixe, L., Baquero, F., Coque, T. M., 2008. Prevalence and spread of extended-spectrum β-lactamase-producing enterobacteriaceae in europe. Clin. Microbiol. Infec. 14 (s1), 144–153.

Collignon, P., Aarestrup, F. M., 2007. Correspondence - extended-spectrum b-lactamases, food, and cephalosporin use in food animals. Clin. Infect. Dis. 44 (10), 1391.

Cooper, B. S., Medley, G. F., Stone, S. P., Kibbler, C. C., Cookson, B. D., Roberts, J. A., Duckworth, G., Lai, R., Ebrahim, S., Wachter, K. W., 2004. Methicillin-Resistant Staphylococcus aureus in hospitals and the community: Stealth dynamics and control catastrophes. Proc. Natl. Acad. Sci. U. S. A. 101 (27), 10223–10228.

Coque, T. M., Oliver, A., Perez-Diaz, J. C., Baquero, F., Canton, R., 2002. Genes encoding tem-4, shv-2, and ctx-m-10 extended-spectrum beta-lactamases are carried by multiple klebsiella pneumoniae clones in a single hospital (madrid, 1989 to 2000). Antimicrob. Agents Chemother. 46 (2), 500–510.

Diekmann, O., Heesterbeek, J. A. P., 2000. Mathematical epidemiology of infectious diseases - model building, analysis and interpretation. Wiley.

Hall, M. A. L.-v., Fluit, A. C., Paauw, A., Box, A. T. A., Brisse, S., Verhoef, J., 2002. Bacteriology - evaluation of the Etest ESBL and the BD phoenix, VITEK 1, and VITEK 2 automated instruments for detection of extended-spectrum beta-lactamases in multiresistant *Escherichia coli* and *Klebsiella* spp. J. Clin. Microbiol. 40 (10), 3703.

Harris, A. D., McGregor, J. C., Johnson, J. A., 2007a. Risk factors for colonization with extended-spectrum β-lactamase–producing bacteria and intensive care unit admission. Emerg. Infect. Dis. 13 (8), 1144–1149.

Harris, A. D., Perencevich, E. N., Johnson, J. K., Paterson, D. L., Morris, J. G., Strauss, S. M., Johnson, J. A., 2007b. Brief reports - patient-to-patient transmission is important in Extended-Spectrum b-Lactamase-producing Klebsiella pneumoniae acquisition. Clin. Infect. Dis. 45 (10), 1347.

Helfand, M. S., Bonomo, R. A., 2006. Extended-spectrum b-lactamases in multidrug-resistant escherichia coli: Changing the therapy for hospital-acquired and community-acquired infections. Clin. Infect. Dis. 43 (11), 1415.

Laupland, K. B., Church, D. L., Vidakovich, J., Mucenski, M., Pitout, J. D. D., 2008. Community-onset extended-spectrum β-lactamase (esbl) producing escherichia coli: Importance of international travel. J. Infect. 57 (6), 441.

Markovska, R., Schneider, I., Keuleyan, E., Sredkova, M., Ivanova, D., Markova, B., Lazarova, G., Dragijeva, E., Savov, E., Haydouchka, I., Hadjieva, N., Setchanova, L., Mitov, I., Bauernfeind, A., 2008. Extended-spectrum β-lactamase-producing enterobacteriaceae in bulgarian hospitals. Microb. Drug Resist. 14 (2), 119–128.

Paterson, D. L., Hujer, K. M., Hujer, A. M., Yeiser, B., Bonomo, M. D., Rice, L. B., Bonomo, R. A., 2003. Extended-spectrum beta-lactamases in klebsiella pneumoniae bloodstream isolates from seven countries: Dominance and widespread prevalence of shv- and ctx-m-type beta-lactamases. Antimicrob. Agents Chemother. 47 (11), 3554–3560.

Pitout, J. D. D., Hanson, N. D., Church, D. L., Laupland, K. B., Jun 2004. Population-based laboratory surveillance for escherichia coli-producing extended-spectrum beta-lactamases: importance of community isolates with blactx-m genes. Clin. Infect. Dis. 38 (12), 1736–1741.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org`

Rodriguez-Bano, J., Navarro, M. D., Romero, L., Muniain, M. A., de Cueto, M., Rios, M. J., Hernandez, J. R., Pascual, A., 2006. Bacteremia due to extended-spectrum beta-lactamase producing escherichia coli in the ctx-m era: A new clinical challenge. Clin. Infect. Dis. 43 (11), 1407–1414.

Romero, E. D. V., Padilla, T. P., Hernández, A. H., Grande, R. P., Vázquez, M. F., García, I. G., García-Rodríguez, J. A., Muñoz Bellido, J. L., 2007. Prevalence of clinical isolates of escherichia coli and klebsiella spp. producing multiple extended-spectrum β-lactamases. Diagn. Microbiol. Infect. Dis. 59 (4), 433–437.

Spall, J. C., 1998. An overview of the simultaneous perturbation method for efficient optimization. J. Hopkins Apl. Tech. D. 19, 482–492.

Stobberingh, E., Arends, J., Hoggkamp-Korstanje, J., Goessens, W., Visser, M., Buiting, A., Debets-Ossenkopp, Y., van Ketel, R., van Ogtrop, M., Sabbe, L., Voorn, G., Winter, H., van Zeijl, J., 1999. Occurrence of extended-spectrum betalactamases (esbl) in dutch hospitals. Infection 27 (6), 348–354.